DTIC FILE COPY

AD-A203 886

ESL-TR-88-16

HAZARD RESPONSE MODELING UNCERTAINTY (A QUANTITATIVE METHOD)

S.R. HANNA, T. MESSIER, L.L. SCHULMAN

SIGMA RESEARCH CORPORATION
394 LOWELL STREET, SUITE 12
LEXINGTON MA 02173

OCTOBER 1988

FINAL REPORT

JULY 1987 - JANUARY 1988

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

AFESC

ENGINEERING & SERVICES LABORATORY
AIR FORCE ENGINEERING & SERVICES CENTER
TYNDALL AIR FORCE BASE, FLORIDA 32403

89 1 30 181

NOTICE

PLEASE DO NOT REQUEST COPIES OF THIS REPORT FROM HQ AFESC/RD (ENGINEERING AND SERVICES LABORATORY). ADDITIONAL COPIES MAY BE PURCHASED FROM:

NATIONAL TECHNICAL INFORMATION SERVICE
5285 PORT ROYAL ROAD
SPRINGFIELD, VIRGINIA 22161

FEDERAL GOVERNMENT AGENCIES AND THEIR CONTRACTORS REGISTERED WITH DEFENSE TECHNICAL INFORMATION CENTER SHOULD DIRECT REQUESTS FOR COPIES OF THIS REPORT TO:

DEFENSE TECHNICAL INFORMATION CENTER
CAMERON STATION
ALEXANDRIA, VIRGINIA 22314

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release Distribution Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) ESL-TR-88-16
6a. NAME OF PERFORMING ORGANIZATION Sigma Research Corporation / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Engineering and Services Center
6c. ADDRESS (City, State, and ZIP Code) 394 Lowell Street, Suite 12 Lexington, MA 02173	7b. ADDRESS (City, State, and ZIP Code) H.Q. AFESC/RDVS Tyndall Air Force Base, FL 32403
8a. NAME OF FUNDING/SPONSORING ORGANIZATION / 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F08635-87-C-0367
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
65502F	3005	00	17

11. TITLE (Include Security Classification)
Hazard Response Modeling Uncertainty (Unclassified)
(A Quantitative Method)

12. PERSONAL AUTHOR(S)
Hanna, S.R., Messier, T., Schulman, L.L.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 87 07 15 TO 88 01 15	October 1988	

16. SUPPLEMENTARY NOTATION
Availability of this report is specified on reverse of front cover.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Toxic Hazards Uncertainty Hazard Response Modeling Dispersion Modeling Meteorology. (AU)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

There are currently available many microcomputer-based hazard response models for calculating concentrations of hazardous chemicals in the atmosphere. The uncertainties associated with these models are not well-known and they have not been adequately evaluated and compared using statistical procedures where confidence limits are determined. The U.S. Air Force has a need for an objective method for evaluating these models, and this project provides a framework for performing these analyses and estimating the model uncertainties. As part of this research, available models and data sets were collected, methods for estimating uncertainties due to data input errors and stochastic effects were developed, a framework for model evaluation was put together, and preliminary applications using test data sets took place. Keywords: Air pollution, Gas dispersion, Toxic hazards.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Captain Lawrence Key	22b. TELEPHONE (Include Area Code) 904-283-4234 / 22c. OFFICE SYMBOL HQ AFESC/RDVS

DD Form 1473, JUN 86

Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

(The reverse of this page is blank)

PREFACE

This report was prepared by Sigma Research Corporation, 394 Lowell Street, Suite 12, Lexington, Massachusetts 02173, under the Small Business Innovative Research (SBIR) Phase I program, Contract Number F08635-87-C-0367, for the Air Force Engineering and Services Center, Engineering and Services Laboratory (AFESC/RDVS), Tyndall Air Force Base, Florida 32403.

This report summarizes work done between 15 July 1987 and 15 January 1988. AFESC/RDVS project officer was Captain Lawrence E. Key.

Because this is an SBIR report, it is being published in the same format in which it was submitted.

This report has been reviewed by the Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

LAWRENCE E. KEY, Capt, USAF
Project Officer

KENNETH T. DENBLEYKER, Maj, USAF
Chief, Environmental Sciences Branch

THOMAS J. WALKER, Lt Col, USAF, BSC
Chief, Environics Division

LAWRENCE D. HOKANSON, Colonel, USAF
Director, Engineering and Services Laboratory

Accession For		
NTIS GRA&I		X
DTIC TAB		☐
Unannounced		☐
Justification		
By		
Distribution/		
Availability Codes		
Dist	Avail and/or Special	
A-1		

(The reverse of this page is blank.)

TABLE OF CONTENTS

TABLE OF CONTENTS
(CONCLUDED)

APPENDIX

LIST OF FIGURES

LIST OF FIGURES
(CONCLUDED)

LIST OF TABLES

LIST OF TABLES
(CONCLUDED)

SECTION I

INTRODUCTION

There are no standard objective quantitative means of evaluating currently available microcomputer-based hazard response models. A number of such models have been recently proposed, and many of them include up-to-date algorithms on important scientific phenomena such as evaporative emissions, dense-gas slumping, and transition to nonbuoyant dispersion. A few data sets exist for testing these models but they have not been tested or compared with the data on the basis of standard statistical significance tests. A review of current vapor cloud models, field data sets and some examples of hazard response model evaluations is given by Hanna and Drivas (Reference 1).

The U.S. Air Force, among others, has increased emphasis on calculating "toxic corridors" due to potential release of hazardous chemicals. The Ocean Breeze/Dry Gulch (OB/DG) model was originally used for calculating these corridors, and contains an estimate of model uncertainty. However, the OB/DG model does not account for dense gas slumping or transient releases. Kunkel (Reference 2) has developed an improved model (AFTOX) that accounts for many of these phenomena (but not dense gases). The generation of models is more advanced scientifically than the OB/DG model, but the new models do not account for uncertainty. The Phase I research described below leads to a preliminary quantitative means of assessing this uncertainty and evaluating these models.

The Phase I research is intended to determine the feasibility of the research program, which then may be carried out in a comprehensive fashion in Phase II. In this case, the Phase I research has had the objectives of reviewing the literature on hazardous response modeling uncertainty, developing a framework that accounts for the three components of the uncertainty (model physics errors, data input errors, and stochastic fluctuations), and applying the preliminary procedures to several models using data sets such as the Prairie Grass data, the Ocean Breeze/Dry Gulch data, the Green Glow data, and the Thorney Island data. The research has attempted to answer the following questions:

- Do suitable data sets exist for use in evaluating hazardous response models?

- What are the errors in the data used for input to models?

- Is it possible to obtain a number of current models for evaluation purposes?

- Can a model evaluation framework be developed that accounts for all the components of model uncertainty, including stochastic fluctuations?

- Can the models properly handle the effects of sampling and averaging times and distances of concentration measurements?

- What are the confidence bounds on model evaluation statistics such as the mean square error? Are they small enough to permit the relative performance of two or more models to be distinguished?

The preliminary study conducted as Phase I of this research plan has resulted in a set of conclusions and recommendations concerning the applicability of the methods developed to quantify the uncertainty in hazardous response modeling. To arrive at this result, the following work tasks have been completed.

Task 1: Literature Review; Collection of Models and Data Sets.
Task 2: Study Components of Model Uncertainty.
Task 3: Develop Framework of Model Evaluation Procedure.
Task 4: Perform Preliminary Application of Procedure.

The results of these tasks are given in Sections II through V, and conclusions and recommendations are given in Section VI.

SECTION II

LITERATURE REVIEW

The first task in any research program is a review of the literature, so that all relevant work by other researchers is considered. This review has included the collection of literature describing appropriate hazardous response models and data sets. In the case of the Air Force Toxic (AFTOX) model systems (Reference 2) the model and the test data sets were obtained directly from the author. Other reviews include those by Kunkel (References 3 and 4), Carney (Reference 5), Ermak and Merry (Reference 6) and Hanna and Drivas (Reference 1). In addition, Spicer and Havens (Reference 7) have evaluated three models with USAF/N_2O_4 test data for the Air Force Engineering and Service Center (AFESC), Key and Bowman (Reference 8) have developed and tested the HARM model for hazardous response modeling, and other DOD groups have been developing similar models (for example, the HAZZARD model at Dugway Proving Ground and the D2PC model (Reference 9) at Aberdeen Proving Ground). Table 1 summarizes some additional hazard response model evaluation exercises found in the literature. More detailed reviews of these references and of available models and data sets are provided below.

A. REVIEW OF SPECIFIC USAF HAZARD MODELING EXPERIENCE

Because the U.S. Air Force handles many toxic chemicals it needs to estimate the atmospheric impact of releases of such chemicals. The 25 years of USAF research on this topic are reviewed below.

1. OB/DG Model Developments

The Ocean Breeze/Dry Gulch (OB/DG) model (Reference 27) was developed for use in support of rocket fuel handling operations at Cape Canaveral and Vandenberg. Dispersion data were collected at those two sites (Cape Canaveral, Florida = Ocean Breeze; Vandenberg AFB, California = Dry Gulch) and at the Prairie Grass, Kansas, site during the 1950s and 1960s (References 28 and 29). These data were used to develop a purely empirical correlation known as the OB/DG model:

TABLE 1. SOME EXAMPLES OF HAZARDOUS GAS DISPERSION MODEL EVALUATIONS

Authors	Models	Data Sets
Layland et al. (Reference 10)	INPUFF 2.0, DEGADIS, OME, PUFF	Eagle (N_2O_4), Thorney Island (Freon)
Paine et al. (Reference 11) Heinold et al. (Reference 12)	AIRTOX	Frenchman Flat (NH_3), Thorney Island (Freon), Burro (LNG), Coyote (LNG)
Ermak et al. (Reference 13) Ermak and Chan (Reference 14)	Gaussian, SLAB FEM3	Burro (LNG), Eagle (N_2O_4)
McRae (Reference 15)	OB/DG, Gaussian	Eagle (N_2O_4)
Alp et al. (Reference 16)	COBRA, HEGADAS	Maplin Sands (LNG)
Riou and Saab (Reference 17)	Box, MERCURE-GL	Thorney Island (Freon)
Balentine and Eltgroth (Reference 18)	CHARM	Burro (Lgn), Eagle (N_2O_4)
Lewellen et al. (Reference 19)	MESO models, Gaussian, ADPIC, IMPACT, others	INTEL SF_6 data
Wheatley et al. (References 20 and 21)	Picknett, DENZ	Thorney Island (Freon)
Puttock and Colenbrander (References 22 and 23)	HEGADIS	Maplin Sands (LNG), Thorney Island (Freon)
Fay and Ranck (Reference 24)	Their own model	Porton (Freon), van Ulden data, Thorney Island (Freon)
Havens and Spicer (Reference 25)	DEGADIS	Burro (LNG), Maplin Sands (LNG), Thorney Island (Freon), Welker (LPG)
Spicer and Havens (Reference 26)	DEGADIS, OB/DG, Gaussian	Eagle (N_2O_4)

$$C_p/Q = 0.00211\ x^{-1.96}\ \sigma_\theta^{\ -0.506}\ (\Delta T + 10)^{4.33} \tag{1}$$

$$\text{or} \quad C_p/Q = 0.000175\ x^{-1.95}\ (\Delta T + 10)^{4.92} \tag{2}$$

where the ratio of the concentration to the source strength C_p/Q is in s m^{-3}, the downwind distance x is in m, the standard deviation of wind direction fluctuations σ_θ is in deg, and ΔT is defined as the temperature difference (^oF) between the 54 ft. and 6 ft. levels on a tower. Wind speed is absent because it is strongly correlated with ΔT. Equation (2) accounts for the strong correlation between σ_θ and ΔT. Stabilities ranged from neutral to unstable during most of these tests.

Predictions of Equation (1) are compared with 321 observations in Figure 1, showing that 72 percent of the predictions are within a factor of two of the observations, and 97 percent are within a factor of four (Reference 27). The OB/DG model was derived using a special subset of data taken from the Ocean Breeze and Dry Gulch experiments. Another subset of data from the same experiments is used in Figure 1. This information on model variability is used by the OB/DG equation to build uncertainty into the model. For example, the model predicts the "toxic corridor length," or distance from the source until a certain concentration is reached. Based on the variability discussed above, the "95th percentile" toxic corridor length is about two times the "50th percentile" or median length. The meaning of "95th percentile" length is the length such that 95 percent of observations would be expected to be less than that value for a given set of input parameters.

2. Evaluations of OB/DG Model

The U.S. Air Force Scientific Advisory Board Ad Hoc Committee on Dispersion of Denser than Air Gases recommended on 9 November 1983 that the OB/DG model be evaluated and possibly replaced with a current state-of-the-art dispersion model. The U.S. Air Force supported two specific reviews of the OB/DG model (References 30 and 31). Ohmstede et al. (Reference 30) compared the model predictions with observations at several sites and recommended that

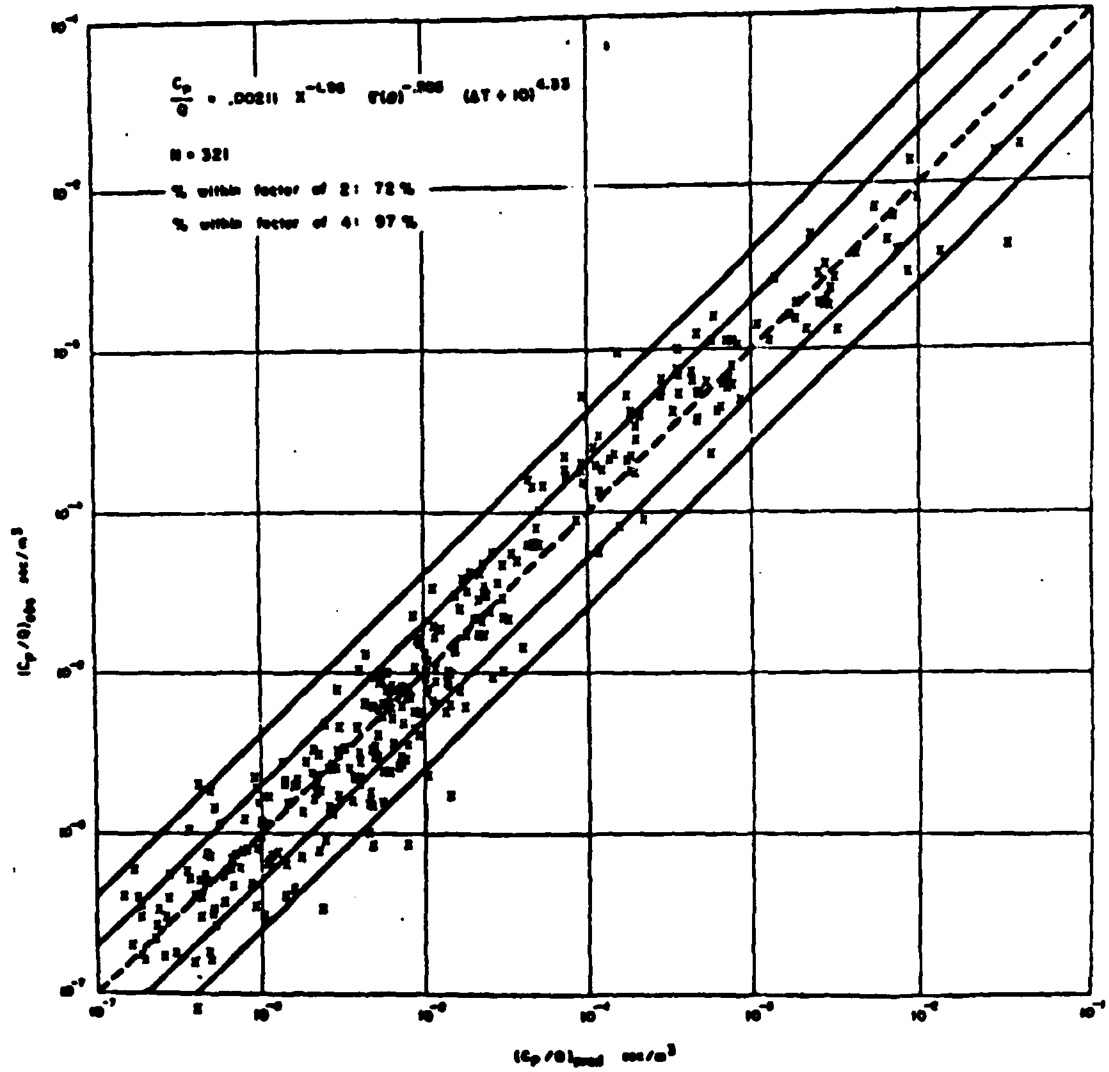


Figure 1. Observed vs. Predicted C_p/Q: Independent Data Test of Final Diffusion Prediction Equation (1) (Reference 9).

the OB/DG model be used only for small releases during the daytime at downwind distances less than about 5 km. They point out that the model is useful only for near-surface continuous point source releases of neutrally buoyant gases.

Kunkel (Reference 31) also found problems with the OB/DG model at night, since few nighttime data were included in the derivation of the model. For light-wind stable conditions, the model underpredicts. However, for windy stable conditions, the model overpredicts by a factor of about four.

3. TOXCOR Model

The Air Weather Service modified the OB/DG model for application to a wide variety of chemical sources. Clewell (Reference 32) and Kahler et al. (Reference 33) describe the TOXCOR model, which is basically the OB/DG model with semi empirical corrections for the molecular weight and vapor pressure of the chemical. Hydrazine, monomethyl hydrazine and N_2O_4 are included in this model.

4. AFTOX Model Development

Following the 9 November 1983 recommendation of the Scientific Advisory Board Ad Hoc Committee on Dispersion of Denser than Air Gases, Bruce Kunkel of AFGL began development of a dispersion model based more on the current state of the art. The resulting model, called AFTOX, is described in a user's guide prepared by Kunkel (Reference 2), who includes test cases and comparisons with the OB/DG model. This development began with an evaluation of existing spill evaporation models (Reference 3) and transport and dispersion models (Reference 31). The latter report considered the OB/DG, the Shell SPILLS model and a modified Shell SPILLS model. The modification consists mainly of the addition of an algorithm from Smith (Reference 34) that permits continuous stability classes to be used. The resulting predictions are much smoother functions, as seen in Figure 2. This modified Shell model evolved into the AFTOX model, which has a Gaussian basis and can be applied to both continuous and instantaneous sources. Because it does not apply to dense gas sources, its application is intended for Air Force sites without the potential for large releases of dense gases.

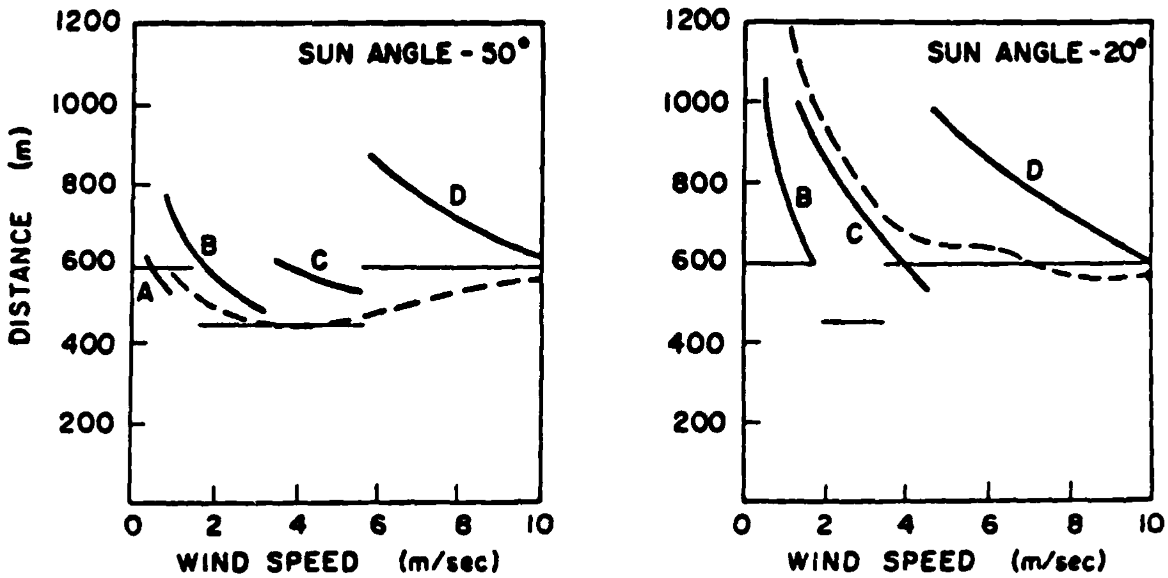


Figure 2. Model Estimates of the Hazard Distance for Benzene as a Function of Wind Speed for a High (50°) and Low (20°) Sun Elevation Angle. The source strength is 1 kg/sec and the concentration of interest is 30 mg/m^3. The light solid line represents the OB/DG model, the heavy solid line represents the Shell Model, and the dashed line represents the Modified Shell Model. The letters represent the Pasquill stability category used in the Shell Model (Reference 13).

The report by Kunkel (Reference 2) compares the AFTOX model with the OB/DG model for the Ocean Breeze, Dry Gulch, Prairie Grass, and Green Glow field studies (these same data are evaluated later in Section V). The AFTOX model is shown to fit the Ocean Breeze, Dry Gulch and Prairie Grass data as well as the OB/DG model (which was derived from these same data), and performs better at the Green Glow site. The figures presented in Reference 2 show the typical AFTOX model error. For example, Figure 3 shows that, at any given observed C/Q, there is a range of about an order of magnitude in the predicted C/Q.

5. Models for Sites where Liquid Propellant may be Spilled.

The OB/DG model was only of limited value at many sites where large amounts of liquid propellant were stored. A few accidental spills occurred, and the resulting plumes were observed to have a variety of behaviors, ranging from dense gas slumping to cases where the cloud rose up several hundred meters in the atmosphere. This behavior depended on the initial source conditions. The AFESC supported an adaptation of the CHARM model to cases where dense gas slumping may occur. RADIAN (Reference 35) describes how the CHARM model was updated to include N_2O_4 and Aerozine-50 and Balentine and Eltgroth (Reference 18) evaluate the revised CHARM model with N_2O_4 and LNG field experiments. They compare the CHARM predictions with the predictions of three Gaussian models, finding that the Gaussian model σ_z's are a factor of ten times the CHARM σ_z's, and that all Gaussian model concentration predictions are below the observations.

Independently, Key and Bowman (Reference 8) reported on Hypergolic Accidental Release Model (HARM) for application to accidental spills at Titan II sites. A rocket exhaust diffusion model developed by the H.E. Cramer Co. was used as a basis for HARM, which includes special provisions for hypergolic reactions and the resulting rise of a buoyant puff into the atmosphere:

$$\text{Puff Rise } z_1 = \left[\frac{8F_I}{\gamma^3 s} + \left(\frac{r_o}{\gamma}\right)^4 \right]^{1/4} - \frac{r_o}{\gamma} \qquad (3)$$

PREDICTED vs. OBSERVED C/Q

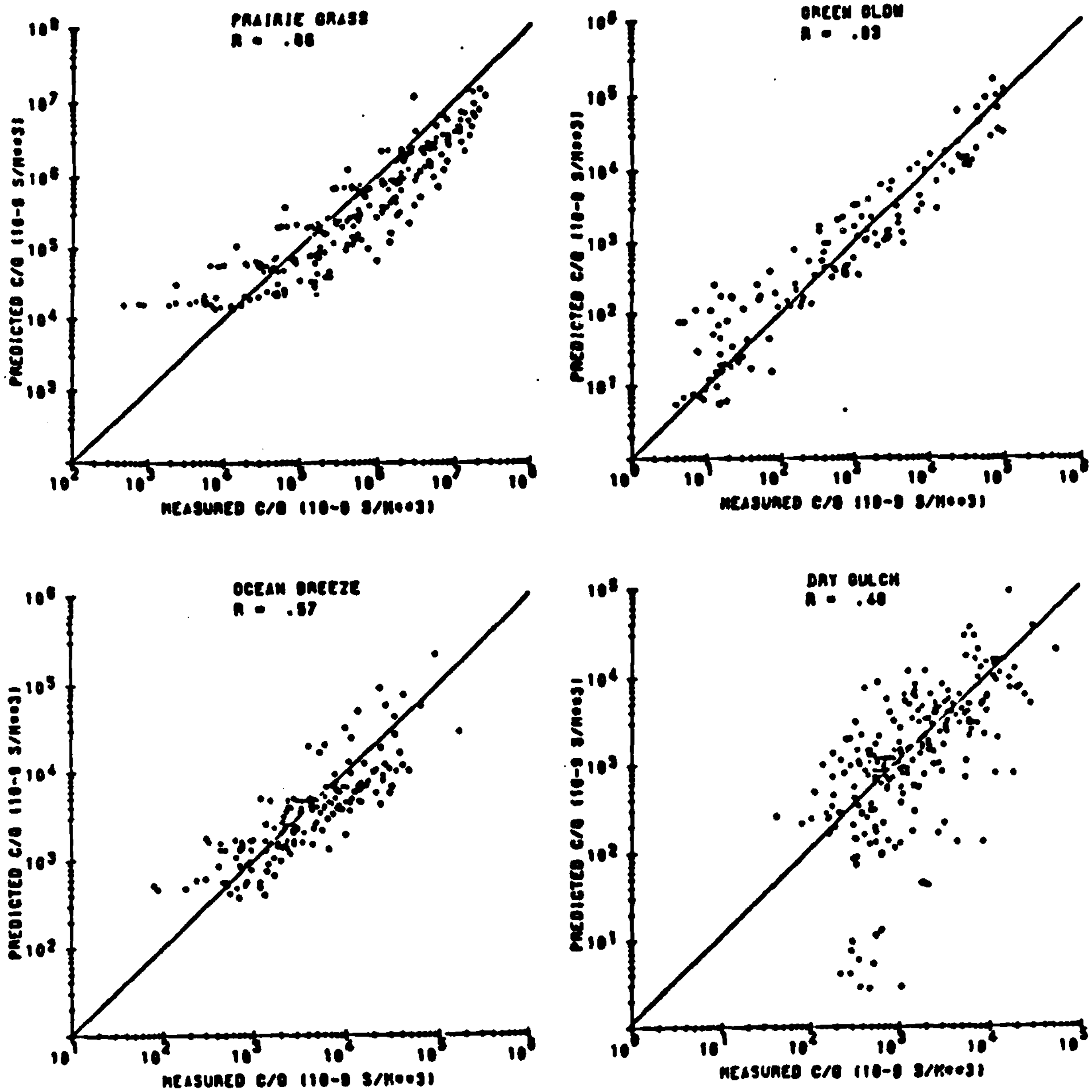


Figure 3. Predictions of the AFTOX Model Plotted versus Observations (from Reference 2).

where γ = 0.4 (entrainment constant)

s = $(g/T)\,(d\theta/dz)$ (stability parameter)

F_I = $3gH/(4\pi c_p \rho T)$ (heat flux)

r_o = initial radius

The parameter g is the acceleration of gravity (9.8 m/s^2), T is ambient temperature (^oK), θ is potential temperature (^oK), z is height (m), C_p is specific heat of air at constant pressure ($J\ gm^{-1}\,{}^oK^{-1}$), and ρ is air density ($gm\ m^{-3}$). The parameter H is the instantaneous heat release in joules. The instantaneous buoyancy flux, F_I, is produced by the release of heat occurring during the reaction of the rocket propellants.

6. N_2O_4 Field tests

Because little information was available on the dispersion of the liquid propellant N_2O_4 in the atmosphere, the AFESC sponsored a series of six field tests at the DOE facility near Las Vegas. Known as the Eagle experiments, the six field tests involved spills of three to five m^3 of N_2O_4. McRae (Reference 15) describes the field data and presents the results of evaluations of the OB/DG, Shell, CHARM, and Gaussian models. He suggests that a major problem is the correct estimation of the evaporative source term, since the heat balance of the spill depends on the evaporative cooling term.

Of the models tested, only the CHARM model could properly simulate the dense gas slumping. Because the other three models do not have dense gas components, they overestimate σ_z by an order of magnitude and underpredict the concentrations by a factor of two to ten. However, the CHARM model is found to "switch over" too soon to a neutral buoyancy algorithm, for the observed cloud exhibited slumping to relatively large distances. McRae (Reference 15) points out that it is difficult to compare predictions with observations, since the required averaging times are not usually matched.

7. DEGADIS Modification for N_2O_4

The DEGADIS model was developed by the U.S. Coast Guard for application to simulating dense gas sources such as LNG and LPG. The model has no source algorithm and treats only surface-level releases. The U.S. Air

Force sponsored an evaluation of the DEGADIS model with their Eagle N_2O_4 field data (Reference 26). The DEGADIS model, a Gaussian model, and the OB/DG model were compared with the field data, with the results shown in Table 2. The DEGADIS concentration predictions are clearly in the range of the observations, while the other model predictions are an order of magnitude low. However, the source term had to be adjusted to account for the interaction of N_2O_4 with moisture in the air.

Further comparisons of the DEGADIS model with dense gas data from Thorney Island, Maplin Sands, Burro, and Welker experiments are reported in Reference 7. The model predicts the average concentration well, but sometimes underpredicts the absolute maximum concentration by a factor of two to five.

The AFESC then supported an extension of DEGADIS to aerosol releases (i.e., two-phase jets). The revised model was tested with anhydrous ammonia data (the so-called Desert Tortoise experiments) by Spicer, Havens and Key (Reference 36). Figure 4 shows that the DEGADIS predictions are within a factor of two of the data points at distances of 500 and 3000 meters, but are much too high at 100 meters. Because there are only two data points, the significance of these conclusions is very low.

8. Model Sensitivity Studies

During 1986 and 1987, Professor Carney of Florida State University prepared several papers for the AFESC on the sensitivity of the AFTOX, CHARM, and PUFF models to uncertainties in input data (Reference 5). His 1987 paper applied the uncertainty formula suggested by Freeman et al. (Reference 37), which has also been applied by Hanna (Reference 38) to a simplified air quality model. If concentration, C, is an analytical function of the variables x_i (i = 1 to n), then the uncertainty or variance $V_c = \sigma_c^2$ is given by the equation

$$V_c = \sum_{i=1}^{n} \left(\partial C/\partial x_i\right)^2 V_{xi} + \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\partial^2 C/\partial x_i \partial x_i\right)^2 V_{xi} V_{xj} \qquad (4)$$

$$+ \ 0.5 \sum_{i=1}^{n} \left(\partial^2 C/\partial x_i^2\right) V_{xi}^2$$

TABLE 2. COMPARISON OF EAGLE 3 AND EAGLE 6 TEST RESULTS AND GAS DISPERSION MODEL PREDICTIONS (REFERENCE 6).

	Maximum NO_2 Concentration Range (ppm)*	σy (m)	σz (m)
EAGLE 3			
Test Results	500-1040	35	3.8
OB/DG	68-73	--	--
Gaussian Plume	68-73	60.5	31.9
DEGADIS	880-1170	57.6-60.7**	2.3-2.9***
EAGLE 6			
Test Results	160-340	35	7.6
OB/DG	20-21	--	--
Gaussian Plume	25-27	60.5	31.9
DEGADIS	190-220	55.6-56.6**	4.5-4.9***

*The concentration range for the model predictions are for the estimated source evolution rate range.

**calculated as $S_y/\sqrt{2}$

***calculated as $S_z/\sqrt{2}$

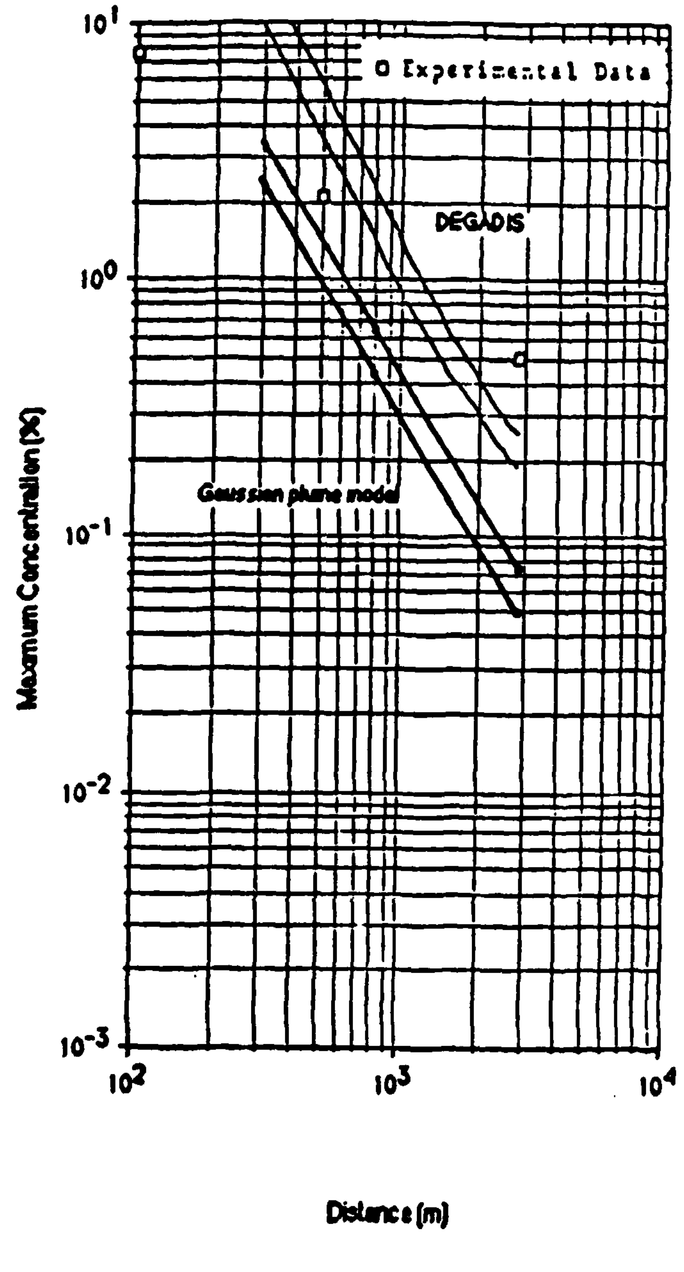


Figure 4. **Maximum Observed Concentration and Maximum Predicted Concentrations Using DEGADIS and the Pasquill-Hanna Gaussian Plume Model for Desert Tortoise 4 (Reference 20).**

where V_{xi} is the uncertainty or variance in input variable x_i. This equation is a Taylor expansion and implicitly assumes that the individual uncertainties are much less than one. Carney (Reference 5) finds that the wind speed, u, contributes the most uncertainty to the concentration, C, predicted by the AFTOX model.

9. New Air Force Model Development: Raj et al. (Reference 39) and Ermak et al. (Reference 40)

Two major new model development efforts are sponsored by the Air Force. The Air Force Geophysics Laboratory (AFGL) and AFESC are supporting the development of an improved dense gas model (ADAM: Air Force Dispersion Assessment Model) that accounts for aerosols and jets (Reference 39) and can be installed at bases where more serious hazardous spills may occur. In addition, AFESC is supporting a modification of the SLAB model (Reference 40) to make it more user-friendly and to allow it to handle transient releases. These new models are expected to be delivered in 1988.

10. Summary of Field Data

Ermak et al. (Reference 40) has put together a comprehensive summary of 26 "bench mark" field experiments, including data from Burro (LNG), Coyote (LNG), Eagle (N_2O_4), Desert Tortoise (NH_3), Maplin Sands (LNG and LPG) and Thorney Island (Freon). This study (funded by AFESC) presents input data required by models and includes observed peak concentrations, average centerline concentrations, and average height and width of the cloud as a function of downwind distance. Presumably these data are sufficiently complete for anyone to run and evaluate his model.

11. Heavy Vapor Model Comparisons

Mercer (Reference 41) describes a model comparison exercise underway in Europe. Several organizations are running their heavy gas models (e.g., DENZ, DEGADIS, DRIFT) on the same data set. Table 3 contains the matrix of input data (25 different cases). Models will be run with a given source emission rate. Unfortunately, there are no baseline data with which to compare the model predictions.

TABLE 3. DATA FOR MERCER (REFERENCE 41) MODEL COMPARISON EXERCISE

MATRIX OF RUNS

		Wind Speed at 10m	1		2	4	8	m/s
		Stability	D	F	D	D	D	
Volume (m^3)	Radius (m)	Roughness length (m)						
2000	7	0.01	x	x	x	x	x	
		0.3	x	x	x	x	x	
	24	0.01	x	x	x	x	x	
$2 \cdot 5 \times 10^5$	120	0.05	x	x	x	x	x	
		1.5	x	x	x	x	x	

Total 25 Cases

It is planned that ADAM (Reference 39) be included in this model comparison.

12. A Methodology for Evaluating Heavy Gas Dispersion Models

In a recent draft report prepared for AFESC, Ermak and Merry (Reference 6) review methods for evaluating heavy gas dispersion models. They first list several specific criteria of interest to the Air Force:

- The methodology is to be based on comparison of model predictions with field-scale experimental observations.
- The methods of comparison must be quantitative and statistical in nature.
- The methods must help identify limitations of the models and levels of confidence.
- The methodology must be compatible with atmospheric dispersion models of interest to the Air Force.

Because these criteria are similar to thos stated for our present study, the results presented by Ermak and Merry are of great use to our research. Their report was received after our Phase I work was completed, and only a brief review was possible. A more extensive study of their results will take place in Phase II.

The Ermak and Merry (Reference 6) report is a review of general evaluation methods and heavy gas model data sets, and does not contain examples of applications of any new evaluation methods with field data sets. Presumably these applications will take place in a later phase of their work. They first review the general philosophy of model evaluation, pointing out that sometimes evaluations of model physics are just as important as quantitative statistic evaluations. Much of their philosophical discussion follows the points made in a review paper by Venkatram (Reference 42). For example, a model whose predictions agree with field data but which contains an irrational physical assumption (e.g. dense gas plumes accelerate upward) is not a good model. Also, they recognize that most model predictions represent ensemble averages, whereas field experiments represent only a single realization of the countless data that make up an ensemble. They emphasize that observed concentrations are strong functions of averaging time, and that most heavy gas dispersion models do not include the effects of averaging time.

Heavy gas dispersion models are distinguished from other dispersion models by three effects: reduced turbulent mixing, gravity spreading, and lingering. The main parameters of interest in evaluations of these models are the maximum concentration, the average concentration over the cloud, and the cloud width and height (all as a function of downwind distance, x). Ermak and Merry emphasize the ratio of predicted to observed variables and define several statistics, such as the mean and the variance. Methods of estimating confidence limits on these statistics are suggested, and the report closes with an example of the application of some of their suggested procedures to a concorted data set draws from a Gaussian distribution.

B. REVIEW OF GENERAL HAZARD MODEL EVALUATION/UNCERTAINTY STUDIES

Section II.A covered specific U.S. Air Force hazard model development and evaluation experience. The present section is intended to cover more general (that is, non-USAF) hazard model evaluation/uncertainty studies. Hanna and Drivas (Reference 1) review several such studies, and Table 1 (presented earlier) contains a list of 13 references on this subject. Most papers and reports on hazard model evaluation deal only with one particular model or data set. Exceptions are papers by Mercer (Reference 43), McNaughton et al. (References 44 and 45) and Layland et al. (Reference 10), which are reviewed first.

1. Comprehensive Model Evaluation Studies

Mercer's (Reference 43) review emphasizes estimation of variability or uncertainty in model predictions, which he finds is typically an order of magnitude when outliers are considered. He includes the following quote from Lamb (Reference 46), which is also appropriate for our discussion.

"The predictions even of a perfect model cannot be expected to agree with observations at all locations. Consequently, the common goal of model validation should be one of determining whether observed concentrations fall within the interval indicated by the model with the frequency indicated, and if not, whether the failure is attributable to sampling fluctuations or is due to the failure of the hypotheses on which the model is based. From the standpoint of regulatory needs the utility of a model is measured partly by the width of the interval in which a majority of observations can be expected to fall. If the width of the interval is very large, the model may provide no more information than one could gather simply by guessing the expected concentration. In particular, when the width of the interval of probable concentration values exceeds the allowable error bounds on the model's predictions, the model is of no value in that particular application."

Mercer (Reference 43) then produces concentration predictions of ten different models for a dense gas source equivalent to that used in the Thorney Island experiments. This comparison (Figure 5) shows that the 10 model predictions range over an order of magnitude at any given downwind distance.

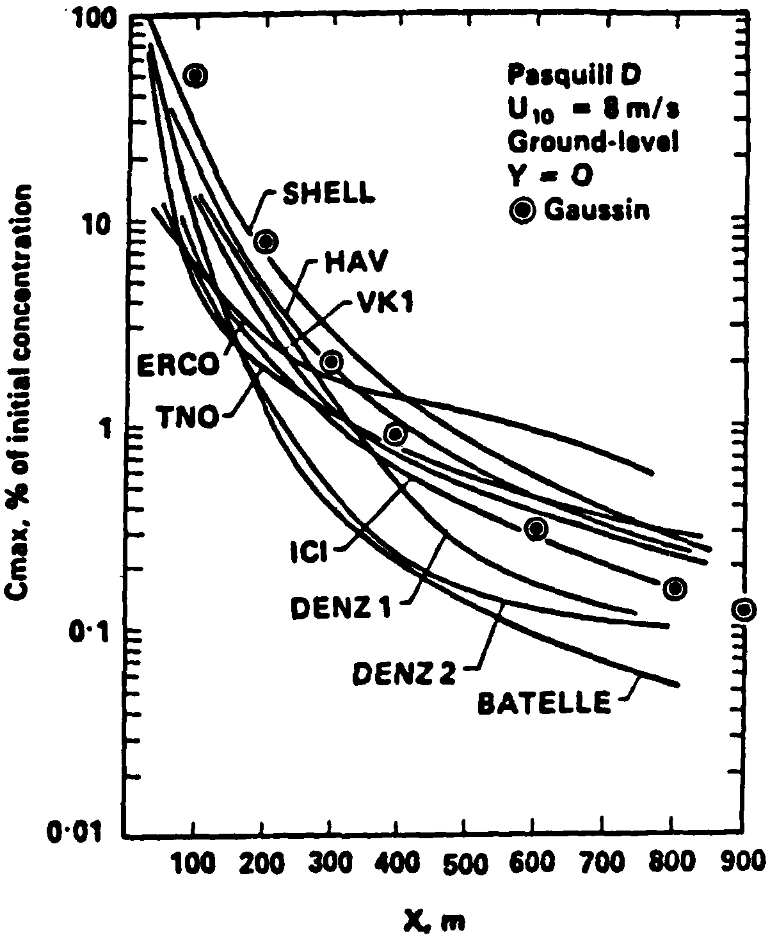


Figure 5. Model Predictions for Thorney Island Trials (Reference 43).

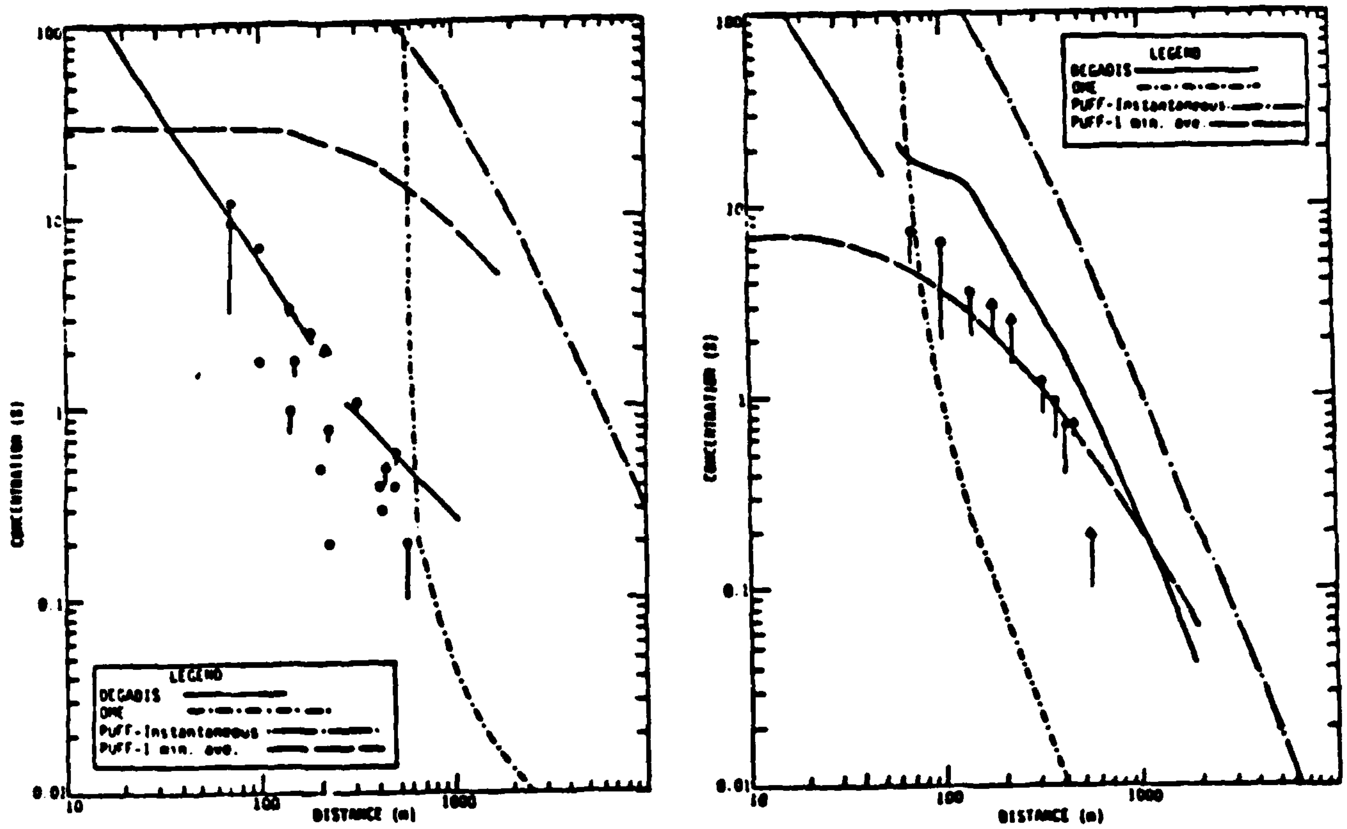
The Chemical Manufacturers' Association (CMA) sponsored an evaluation of eight dense gas dispersion models and nine spill evaporation models (References 44 and 45). The authors ran some of the models themselves and requested the developers of proprietary models to run their own models using standard input data sets. Model uncertainty is typically a factor of two to five. The comparisons are clouded by the use of some data sets that had already been used to "tune" certain of the models tested.

Some of the same authors were involved in a similar study performed for the EPA (Reference 10), in which the DEGADIS, OME, and INPUFF models were compared with some Thorney Island observations. The DEGADIS and OME models account for dense gas slumping. Figure 6 contains the results of these comparisons, which do not give one great confidence in hazard model predictions. Some of this error may be caused by uncertainties in input data specification.

2. Model Comparisons with Thorney Island Data

The Thorney Island experiment was carried out to test the dense gas slumping component of hazard response models. The experiment is reviewed in more detail in Section II.D, but can be briefly described as an instantaneous ground-level release over a flat surface of about 1000 m^3 of freon gas in the shape of a cylinder. The data were made available quickly and completely to the modeling community and have been used for many model comparisons, some of which are discussed below.

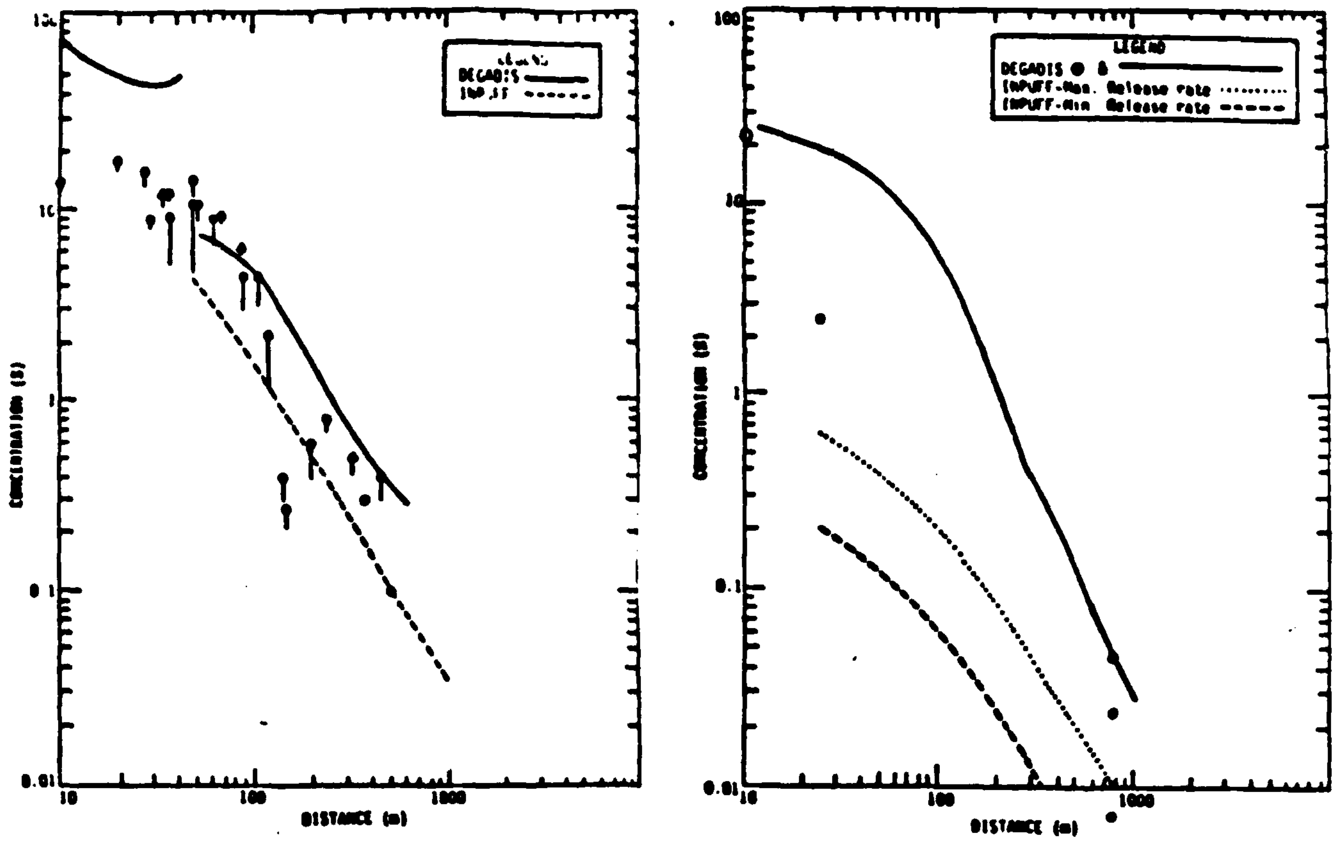
Puttock and Colenbrander (Reference 23) discuss the random nature of the data. They state "it is desirable to know exactly what any model is intended to predict (usually some sort of average), and, in a final assessment, how much one experiment might be expected to deviate from the average." This quote emphasizes the fact that a model prediction represents an ensemble average, which would be given in the atmosphere by an average over 100 or more experiments conducted with the same external conditions. They present Figure 7, which demonstrates the uncertainty in observed concentrations in a single Thorney Island experiment. Data from three different sensor heights are plotted.



Thorney Island Trial 9.

Thorney Island Trial 13.

Figure 6. Model Comparisons Published in Reference 10.



Thorney Island Trial 45.

Eagle Trial 6.

Figure 6. Model Comparisons Published in Reference 10 (Concluded).

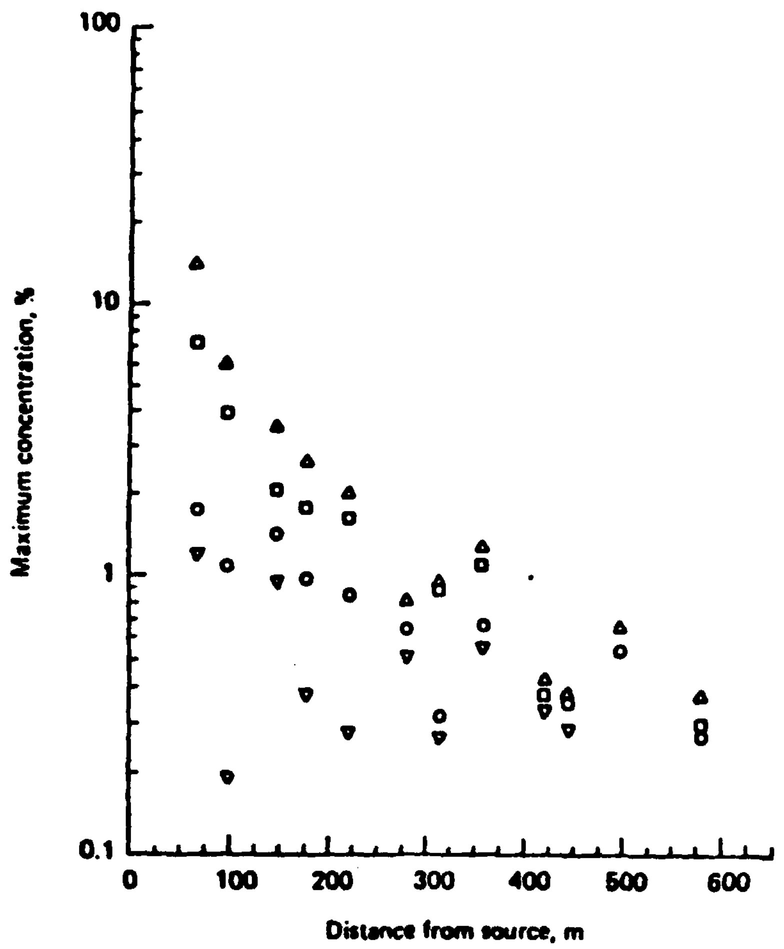


Figure 7. The Maximum Gas Concentration Observed at Various Locations in Thorney Island Trial 7, Plotted Against Distance from the Source (Reference 23).

Puttock (Reference 47) later compares the predictions of his HEGABOX/HEGADAS model with observations during 13 Thorney Island trials. Table 4 presents the observed and predicted distances where peak concentrations equal 5 percent, 2 1/2 percent and 1 percent. Generally, the observations and predictions are within ± 50 percent. He finds that the roughness length used in the model should be based on upwind terrain and therefore on wind direction.

The uncertainties in models were also emphasized by Wheatley et al. (Reference 48), who used the Thorney Island data to derive the top entrainment constant necessary to make models such as DENZ agree with the observations. Figure 8 presents the range of derived top entrainment constants for 13 Thorney Island trials. The symbols S_2 and S_5 refer to goodness-of-fit measures employed in the maximum likelihood procedure. The range is typically an order of magnitude.

The American Petroleum Institute (API) supported a comparison by EAI (Reference 49) in which the Eidsvik, MARIAH II, HEGADAS II, and Cox and Carpenter models were evaluated with the Thorney Island data. These four models were chosen because they represent the three model groups:

Box models: Eidsvik; Cox and Carpenter
K-theory model: HEGADAS-II
Hydrodynamic 3D model: MARIAH-II

This report is very useful because it contains complete sets of tables and figures that can be used by other researchers for further analysis. It is concluded that the Eidsvik and Cox and Carpenter models are conservative in the sense that they overpredict by a factor of about two, while the HEGADAS and MARIAH models are closer to the observations. Consequently, the first two models are recommended for screening analyses, while the last two models are recommended for more refined calculations. An example of a comparison of the four models with data from Trial 15 is presented in Figure 9. The tendency of the Eidsvik and Cox and Carpenter models to overpredict can be clearly seen.

TABLE 4. THORNEY ISLAND OBSERVATIONS AND HEGABOX/HEGADAS MODEL PREDICTIONS (REFERENCE 47).

Observed and predicted distances for peak concentrations to decay to 5%, 2½% and 1%

Trial number	Volume, m^3	Relative density	Wind speed, m/s	Pasquill stability class	Roughness length, mm	Max. distance to 5%, m		Max. distance to 2½%, m		Max. distance to 1%, m	
						Observed	Predicted	Observed	Predicted	Observed	Predicted
6	1580	1.60	2.7	D	15	100	121	150	187	240	302
7	2000	1.73	3.5	D/E	15	125	123	190	195	≈350	323
8	2000	1.63	2.4	B	11	125	124	≈220	199	≈370	333
9	2000	1.60	1.7	F	8	125	103	195	166	≈360	300
11	2100	1.96	5.1	D	15	110	126	175	199	≈290	328
12	1950	2.37	2.6	F	15	115	101	175	161	290	281
13	1950	2.00	6.5	D	9	130	145	215	224	370	355
14	2000	1.76	6.0	D	2	130	161	215	260	≈400	420
15	2100	1.41	5.4	B	2	180	184	275	290	455	430
16	1580	1.68	4.6	C/D	2½	120	152	205	237	370	335
17	1700	4.20	5.0	D	15	70	112	110	179	185	295
18	1700	1.87	7.4	C/D	4	90	162	150	231	275	330
19	2100	2.12	5.3	D	9	100	135	160	217	≈275	350

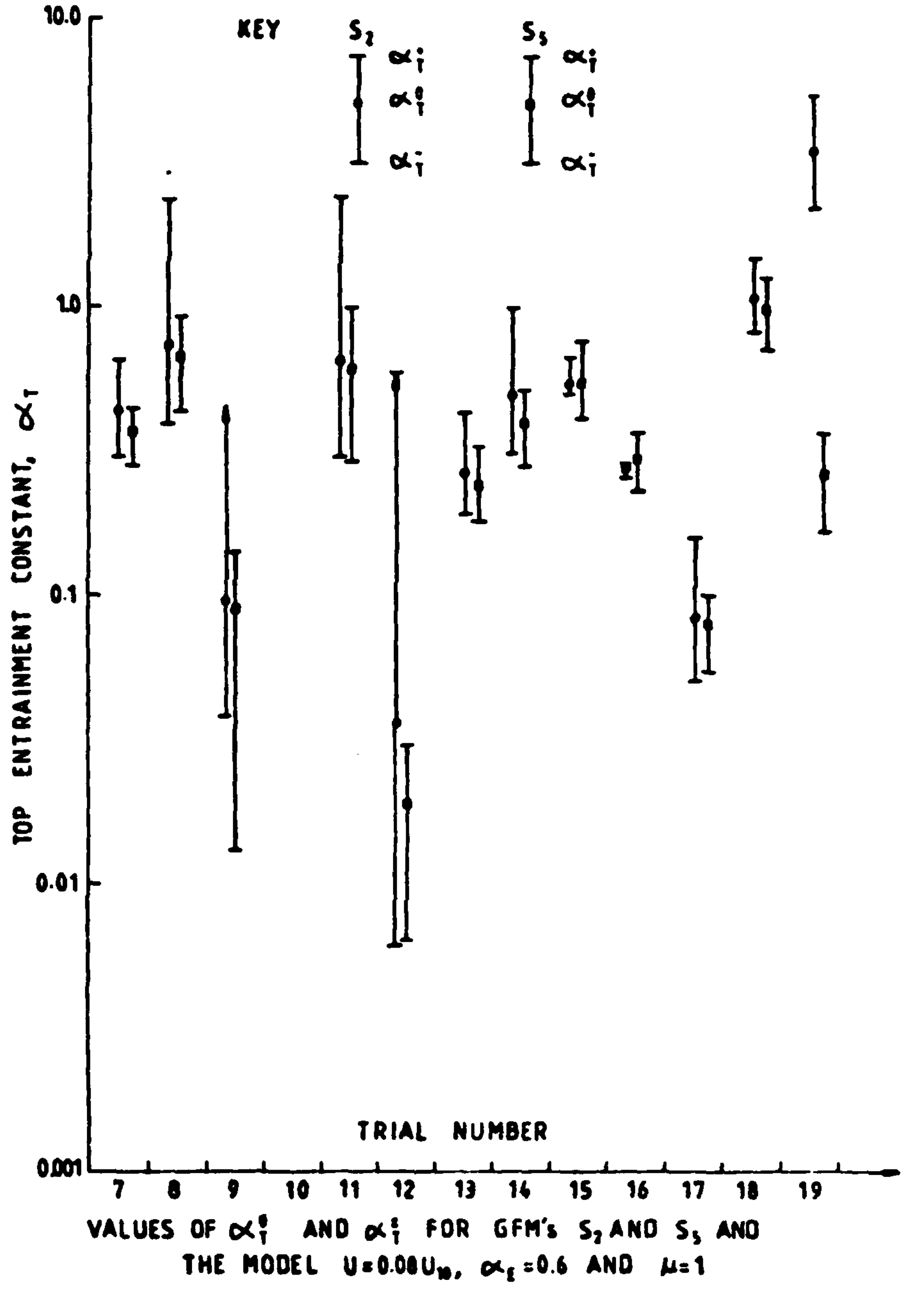


Figure 8. Entrainment Rates Derived from the Thorney Island Data (Reference 48).

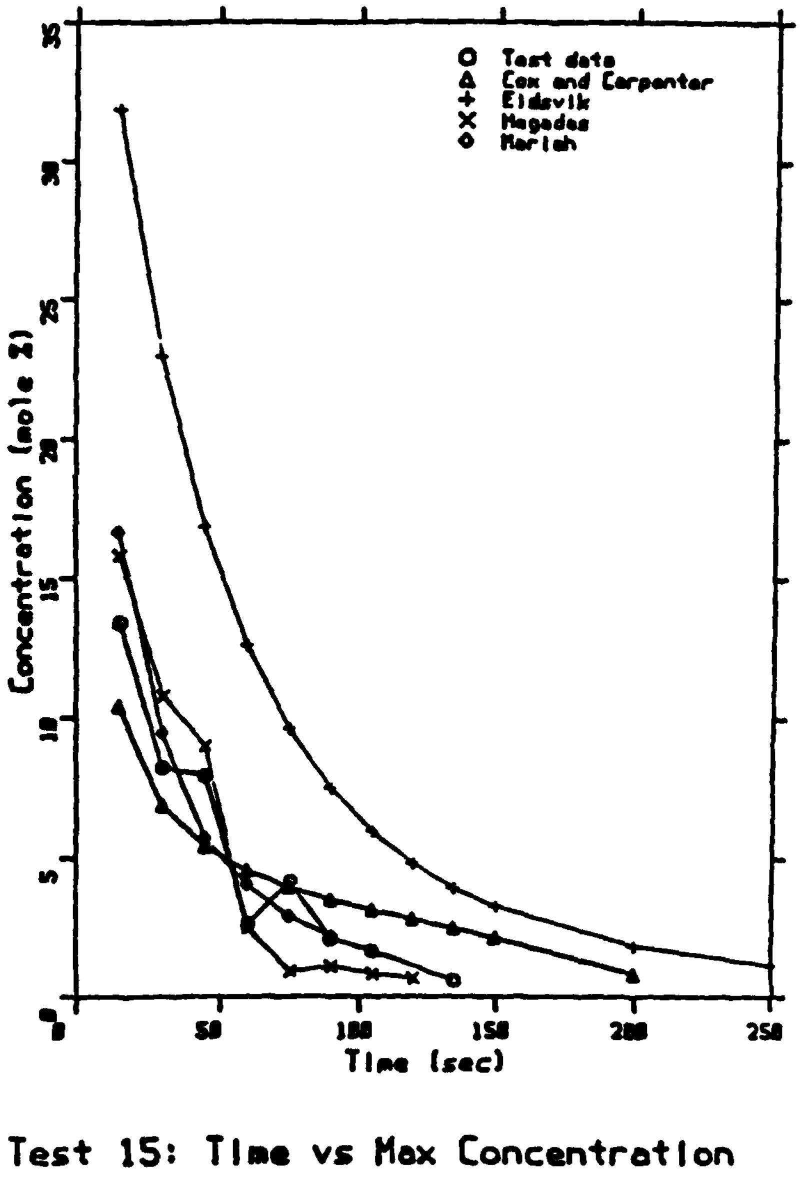


Figure 9. Predictions of Four Models Compared with Thorney Island Test 15 Data (Reference 49).

Our final example of a model evaluation using the Thorney Island data is the paper by Chikhliwala et al. (Reference 50). The so-called SAFER model, whose dense gas slumping algorithm is based on the Kaiser and Walker model, is compared with data from six Thorney Island trials. The agreement appears to be good; however, this may not be an independent comparison, since these same data were used in the development of the model.

3. Model Comparisons with Aerosol Data

The Thorney Island source was highly simplified so that the data could be used to test the dense gas algorithms in models. Many real-world sources consist of two phase jet releases, such as the release of HF, N_2O_4, or NH_3 from a hole in a pressurized tank. Here the major problem is to estimate what fraction of the release consists of gas and what fraction consists of liquid. The situation is further complicated by the fact that part of the liquid may be spilled onto the ground (and subsequently evaporated) and part may be broken up into an aerosol that drifts off downwind. Thus, a release of NH_3, with a molecular weight less than that of air, can act like a dense plume because of the liquid drops within the plume.

Blewitt et al. (Reference 51) discuss applications of the SLAB and DEGADIS models to data collected from series of HF tests. They found that the source term must be adjusted so that all of the HF is assumed to drift off with the gas plume (that is, no liquid spill occurs on the surface). With this adjustment, the SLAB model predictions were within a factor of two of the observations. Analysis of the DEGADIS model predictions was deferred until questions regarding the appropriate averaging time for the model are resolved.

Large-scale spill tests of NH_3 and N_2O_4 at the Nevada Test Site are discussed by Koopman et al. (Reference 52). These tests were named the Desert Tortoise and Eagle experiments (referred to earlier), respectively. Aerosols were found to play a very important role in dense gas dispersion. A simple Gaussian plume model, the empirical OB/DG model, and the three-dimensional hydrodynamic model FEM3 were applied to the data. The Gaussian and OB/DG models were inadequate, while the FEM3 model gave reasonable predictions once the source term was adjusted to account for the presence of aerosols.

C. REVIEW OF AVAILABLE HAZARD RESPONSE MODELS

This section is divided into a relatively short comprehensive review of available hazard response models and a more complete review of available Air Force models.

1. Comprehensive Review of Models

As part of a review for the AIChE, Hanna and Drivas (Reference 1) distributed questionnaires to all known hazard response modelers. This list included developers of publically-available models as well as proprietary models. Because of the absence of a recommended government model, a flourishing business in proprietary hazard response models has grown up over the past few years. A total of 32 completed questionnaires were returned, and the results are tabulated in Table 5. Most of the entries in the table are "Yes" or "No" answers to questions regarding whether or not the model includes a certain component. Both the source emission component and the transport and dispersion component are included. This table represents the status of models as of about January 1987. Since that time new models have appeared and some of the older models have been modified.

Many people are working on specific subcomponents to models. For example, papers on various aspects of dense gas dispersion appear routinely in the scientific journals, but in most cases these models do not find their way into Table 5 because they are not embodied in a specific user-friendly piece of software.

2. Review of Air Force and other Publicly - Available Models

At the start of this project an attempt was made to collect publicly-available hazard response models and to install and test them on microcomputers. Because of the requirement that the models be installed on microcomputers, some models such as FEM3 and DEGADIS were not included (Note: the API is funding an effort to make a PC-compatible version of DEGADIS). At the other extreme, some models such as OB/DG consist of only a single equation and therefore it is not necessary to install it on any computer. Nevertheless, such models are included in our collection. Table 6 contains

(The reverse of this page is blank)

TABLE 5. RESULTS FROM MODEL QUESTIONNAIRES (MOST ANSWERS ARE GIVEN AS Y = YES OR N = NO) AS OF DECEMBER 1986.

Question	AVACTA II	CARE	CHARM	COBRA III	CRUNCH	DEGADIS	DENZ	D2DC	Emissions	EPIDIS	FEM3	GASP	HASTE	EAHAP	Eldsvik
Operating Information															
Form of model: H=Hardware; S=Software	S	H;S	H;S	S	S	S	S	S	S	S	S	S	H;S		H;S
Main use: R=Research; A=Applied	R;A	A	A	A	R;A	R;A	R;A	A	A	A	R	R;A	A	R	R;A
Operate in interactive mode?	N	Y	Y	N	Y	Y	N	Y	N	Y	N	N	Y	Y	Y
Support system?		Y	Y		N	Y	N	Y	N			N	Y	Y	N
Number sold or given away?	2	15	80		>10	>10	>10	40	0				6	5	
Link to emergency system?		Y	Y		N	N	N	N	N			N	Y	N	N
Input Data															
Accept real time weather data?	N	Y	Y	N	N	N	N	N	N	N	N	N	Y		N
Method of data entry: H=Hand; F=data file memory; D=disk or tape	F	H;F;D	H;F;D	H;F	F	H;F	F	H	H;F	H;F;D	F	H;F	H;F		H
Source Emissions Model?	N	Y	Y	Y	N	N	N	Y	Y	Y	Y	Y	Y		Y
Evaporation of Spilled Liquids?		Y	Y	Y				Y			Y	Y	Y	Y	Y
Flashing		Y	Y	Y				N			Y	Y	N	N	Y
Multicomponents		N	N	N				N			Y	N	N	N	N
Entrainment as aerosols?		Y	Y	Y				N			N	Y	N	N	Y
Heat transfer, substrate to cloud?		Y	N	Y				N			Y	Y	Y	Y	Y
Number of substrates (W=Water; S=Soil)		W;S		W;S							W;S	W;S	W;S	W;S	W;S
Mass transfer in liquid phase?		Y	N	N				N			Y	Y	N	Y	Y
Evaporation of aerosols		N	Y	Y				N			N	Y	N	N	Y
Gas flux from container rupture?		N	Y	Y				N			N	Y	N	N	Y
Condens. of moisture in vapor cloud?		N	Y	Y				N			N	Y	Y	N	N
Wind influence on evaporation?		Y	Y	N				Y			Y	Y	N	N	Y
Number of chemicals hardwired (I=input by user)		50,I	83	5,I				17			I	18,I	I	I	I
Transport and Dispersion Model?	Y	Y	Y	Y	Y	Y	Y	Y			N	Y	Y	N	Y
Releases treated: I=Instantaneous; C=Continuous; V=Variable	I;C;V	C;V	I;C;V	V	C	I;C;V	I	I;C;V				V	I;C		I;C;V
Dense cloud?	N	Y	Y	Y	Y	Y	Y	N				Y	Y		Y
Jet?	Y	N	Y	N	N	N	N	Y				N	N		N
Neutral cloud?	Y	Y	Y	N	Y	Y	Y	Y				Y	Y		Y
Buoyant cloud?	Y	N	Y	N	N	N	N	Y				N	N		Y
Surface roughness?	Y	Y	N	Y	Y	Y	Y	Y				Y	Y		Y
Complex terrain handled?	Y	Y	N	N	N	N	N	N				N	Y		N
U variations in time and space?	Y	Y	N	N	N	N	N	N				N	Y		N
Indoor concentrations?	N	N	N	N	N	N	N	N				N	N		N
												N	Y		Y

Buoyant cloud?	Y	N	Y	N	N	N	N	Y		N	N		Y
Surface roughness?	Y	Y	N	Y	Y	Y	Y	Y		Y	Y		Y
Complex terrain handled?	Y	Y	N	N	N	N	N	N		N	Y		N
U variations in time and space?	Y	Y	N	N	N	N	N	N		N	Y		N
Indoor concentrations?	N	N	N	N	N	N	N	N		N	N		N
Building wake effects?	N	Y	Y	N	N	N	N	N		N	Y		Y
Advection/Dispersion Model: B=Box													
or Slab; G=Gaussian; K=K/numerical	G	G	G	B	B;G	K	B;G	G		K	K		B;G
Along-wind dispersion?	Y	Y	Y	Y	N	Y	Y	N		Y	Y		Y
Vertical wind shear?	N	Y	N	Y	Y	Y	Y	N		Y	Y		Y
Chemical reactions in plume?	Y	Y	Y	N	N	Y	N	N		N	N		N
Dry or wet deposition?	Y	Y	N	N	N	N	N	D		N	N		N
Concentration fluctuations?	N	N	N	N	N	N	N	N		Y	N		N
Number of chemicals (I=Input)	I	50,I	83	5,I		5		17		18,I	1		I
Output													
Averaging time (minutes)(I=Input)		15	1-10	3		10				0			
Distance limits (km)								20		70	10		
Evaluation?	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y	N	N
How are data presented													
T=Table; G=Graph		T;G	T;G		T	T	T;G	T;G	T;G	T;G	T;G	T;G	T;G

(The reverse of this page is blank.)

TABLE 5. RESULTS FROM MODEL QUESTIONNAIRES (MOST ANSWERS ARE GIVEN AS Y = YES OR N = NO) AS OF DECEMBER 1986.

HEAVYPUFF	HEGADAS	INPUFF2.0	MIDAS	ModSys	PLUME PATH	RIMPUFF	SAFEMODS	SAFER/TRACE	SLAB	SRI PUFF	TRAUMA	VAPID
S	S	S	S;N	S	S	S	S;N	S;N	S	S	S	S
R	R;A	A	R;A	A	R;A	A	R;A	A	R;A	R	A	R
N	N	N	Y	N	N	Y	Y	Y	N	Y	N	N
N	N	Y	Y	N	N	Y		Y		N	N	
	1	120	28		>10	3		75				
	N	N	Y	N	N			Y			N	
N	N	N	Y	N	N	Y	Y	Y	N	N	N	N
N;F;D	F	D	N;F;D	N;F	F	F	N;F;D	N;F;D	F	N;F;D	F	N;F;D
N	Y	N	Y	Y	N	Y	Y	Y	Y	N	Y	Y
	Y		Y	N			Y	Y	Y		N	Y
	N		Y	N			Y	Y	N		Y	Y
	N		Y	Y		Y	Y	Y	N		N	
	N		N	N			Y	Y	N		N	
	Y		Y	N			Y	Y	Y		N	
	W;S		W;S				W;S	W;S				
	N		Y	N			Y	Y	N		Y	
	N		N	N			Y	Y	N		Y	
	N		Y	Y		Y	Y	Y	N		T	
	Y		N	N			Y	Y	N		Y	
	N		Y	N			Y	Y	N		N	Y
	I		I	2		25	I	I	1		I	
Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
I	C;V	C;V	V	V	C	I;C;V	I;C;V	I;C;V	C	V	C	V
Y	Y	N	Y	Y	Y		Y	Y	Y	N	N	Y
Y	N	N	Y	Y	Y	N	Y	Y	N	N	Y	N
Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Y	N	Y	Y	Y	Y	Y	Y	N	N	N	N	N
Y	Y	N	Y	Y	N	Y	Y	Y	Y	N	N	Y
N	N	N	Y	N	N	Y	Y	Y	N	Y	N	N
N	N	Y	Y	N	N	Y	Y	Y	N	Y	N	N
N	N	N	Y	N	N	N	Y	N	N	N	N	N
N	N	N	Y	N	N	N	N	N	N	N	N	N

Y	Y	N	Y	Y	N	Y	Y	Y	N	Y	N	N
Y	Y	N	Y	Y	N	Y	Y	N	N	N	N	Y
N	N	N	Y	N	N	Y	Y					
N	N	Y	Y	N	N	Y	Y	Y	N	Y	N	N
N	N	N	Y	N	N	N	Y	N	N	N	N	N
N	N	N	Y	N	N	N	N	N	N	N	N	N
B	B;G	G	B;G;K	G;K	G	G	B;G	B;G	B	G	B	B
Y	Y	N	Y	Y	N	Y	Y	Y	N	Y	N	N
Y	Y	N	Y	Y	N	Y	Y	N	N	N	N	Y
N	N	N	N	Y	N	N	Y	Y	N	N	Y	N
N	N	Y	Y	N	N	Y	Y	N	N	N	Y	D
N	N	N	N	Y	N	Y	Y	N	N	N	N	N
4	1		1	2	1	25	900	1	1		1	1
	1	3,60			0		0.5,1		1	60		10
		100	10		1	20	5					10
N	Y	Y	Y	Y		Y	Y	Y	Y		N	N
T;G	T	T	T;G	T;G	T	T;G	T;G	T;G	T;G		T	T;G

(The reverse of this page is blank.)

TABLE 5. RESULTS FROM MODEL QUESTIONNAIRES (MOST ANSWERS ARE GIVEN AS Y = YES OR N = NO) AS OF DECEMBER 1986.

	SAFE	SAFETI	TECJET	WHAZAN
Operating Information				
Form of model: H=Hardware; S=Software	S	S	S	S
Main use: R=Research; A=Applied	A	A	A	A
Operate in interactive mode?	Y	Y	Y	Y
Support system?	Y	Y	Y	Y
Number sold or given away?	1	7	0	60
Link to emergency system?	N	N	N	N
Input Data				
Accept real time weather data?	Y	N	N	N
Method of data entry: H=Hand; F=data file memory; D=disk or tape	H;F;D	H;F;D	H;F	H
Source Emissions Model?	N	Y	Y	Y
Evaporation of Spilled Liquids?		Y	N	Y
Flashing		Y	Y	Y
Multicomponents		N	N	N
Entrainment as aerosols?		Y	Y	Y
Heat transfer, substrate to cloud?		Y	N	Y
Number of substrates (W=Water; S=Soil)		S		S
Mass transfer in liquid phase		Y	N	Y
Evaporation of aerosols		Y	Y	Y
Gas flux from container rupture?		Y	Y	Y
Condens. of moisture in vapor cloud?		Y	Y	Y
Wind influence on evaporation?		N	N	Y
Number of chemicals hardwired (I=input by user)		60,I	62,I	20,I
Transport and Dispersion Model?	Y	Y	Y	Y
Releases treated: I=Instantaneous; C=Continuous; V=Variable	C;V	I;C;V	C	I;C;V
Dense cloud?	N	Y	Y	Y
Jet?	Y	Y	Y	Y
Neutral cloud?	Y	Y	Y	Y
Buoyant cloud?	Y	Y	Y	Y
Surface roughness?	Y	Y	Y	Y
Complex terrain handled?	N	N	N	N
U variations in time and space?	Y	N	N	N
Indoor Concentrations?	N	N	N	N

Surface roughness?	Y	Y	Y	Y
Complex terrain handled?	N	N	N	N
U variations in time and space?	Y	N	N	N
Indoor Concentrations?	N	N	N	N
Building wake effects?	N	N	N	N
Advection/Dispersion Model: B=Box				
or Slab; G=Gaussian; K=K/numerical	G;K	B;G		B;G
Along-wind dispersion?	Y	N	N	N
Vertical wind shear?	Y	N	N	N
Chemical reactions in plume?	Y	N	N	N
Dry or wet deposition?	Y	Y	N	N
Concentration fluctuations?	Y	N	N	N
Number of chemicals (I=Input)	I	60,I	62,I	20,I
Output				
Averaging time (minutes)(I=Input)	60	10	10	10
Distance limits (km)	50	20		
Evaluation?	N	N	Y	N
How are data presented				
T=Table; G=Graph	T;G	T,G	T,G	T,G

(The reverse of this page is blank.)

TABLE 6. HAZARDOUS GAS MODELS INS

	AFTOX	AVACTA II	CHARM	D2PC	MADICT	OB/
REFERENCES	KUNKEL(1987) AFGL TR 132	ZANNETTI(1986) AV-R-86/530	CHARM document version 84.1	WHITACRE(1987) CRDEC-TR-87021	LUDWIG(1985)	AFG
DATA ENTRY	INTERACTIVE	FILE	INTERACTIVE	INTERACTIVE	FILE	FIL
SOURCE MODELS						
• LIQUID JET	NO	NO	YES as in SPILLS	NO	NO	NO
• FLASHING	NO	NO	YES as in SPILLS	NO	NO	NO
• SPILL EVAPORATION	YES same as SPILLS	NO	YES as in SPILLS	YES ORG report 40	NO	NO
• GAS JET	NO	NO	YES	NO	NO	NO
• 2-PHASE JET	NO	NO	YES	NO	NO	NO
DISPERSION MODELS						
• MODEL TYPE	GAUSSIAN	GAUSSIAN	GAUSSIAN	GAUSSIAN	GAUSSIAN	EMP
• INSTANTANEOUS PUFF	YES-gas or liquid	YES gas	YES-gas or liquid	YES gas or liquid	YES gas	NO
• CONTINUOUS PLUME	YES- gas,liq.	YES mult puffs	YES- gas,liq. mult puffs	YES gas or liquid	YES mult. puffs	YES
• DENSE CLOUD	NO	NO	YES	NO	NO	NO
• NEUTRAL CLOUD	YES	YES jet rise	YES	YES jet rise	YES no rise	YES
• BUOYANT CLOUD	YES stack plume	YES	YES	YES plume rise	NO	NO
• NONUNIFORM WIND	NO	YES grid cell winds	NO	NO	YES-preprocsd grid winds	NO
• SURFACE ROUGHNESS	YES adj dispersion	NO	NO	YES- forest effct on disp	NO	NO
• COMPLEX TERRAIN	NO	YES	NO	NO	YES-neutral or stable	NO
• BUILDING WAKES	NO	NO	YES	NO	NO	NO
• CHEMICAL REACTIONS	NO	YES- user supplied %	NO	NO	NO	NO
• DEPOSITION	NO	YES- dry, wet % or dep vel	NO	YES- vapor depl f(u,stab,z0)	NO	NO
OUTPUT						
• DATA FORMAT	graph,file single case	file each time step	graph,file single case	file single case	file	FI
• AVERAGING TIME	user input	user input mst and conc	not specified	dosage model or max conc	60 minutes	10
• DISTANCE LIMITS	> 100 m < 40 km	not specified	not specified	not specified	not specified	5
COMMENTS	64 chemicals user input time duration	version 3.1 user wind fld or wnd fld interpolated	ISC sigma-z wake effect	2-min correctn dosage model spec army depot	wind fields precalc or processed from wind obs	[illegible]

ED LS INSTALLED ON SRC COMPUTERS.

	OB/DG	ONE	PUFF/INPUFF	RVD	SLAB	SPILLS
5)	AFCRL	ONE (1983) ARB-162-83-ARSP	PETERSEN AND LAVDAS(1986)	LAYLAND,et. al. (1986)	ERMAK, et. al. draft July 1987	FLEISCHER(1980) SHELL DEV CO
	FILE	INTERACTIVE	FILE	INTERACTIVE	FILE	INTERACTIVE use tape only
	NO	YES	NO	NO	NO	NO user emis rate
	NO	NO	NO	NO	NO	YES for liq gases during min 1
	NO	YES	NO	NO	YES	YES
	NO	YES	NO	NO	NO	NO user emis rate
	NO	NO	NO	NO	NO	NO
	EMPIRICAL	GAUSSIAN	GAUSSIAN	EMPIRICAL-SIMILARITY	NUMERICAL/BOX crosswind ave-solve cons eqn	GAUSSIAN
	NO	YES-gas or liquid	YES-gas only	NO	NO	YES- gas or liq
fs	YES	YES-gas,liq.	YES-mult puffs	YES	YES	YES- gas,liq mult puffs
	NO	YES-slumping instan. only	NO	YES Ri > 30	YES	NO
	YES	YES	YES	NO Ri <= 30	NO	YES
	NO	NO	YES Briggs	NO	NO	YES stack,area or point
ced b	NO	NO	YES user supplied	NO	NO	NO
	NO	NO	NO	NO but need u*	YES	NO
i	NO	NO	NO	NO	NO	NO
	NO	NO	NO	NO	NO	NO
	NO	NO	NO	NO	NO	NO
	NO	NO	YES dry only	NO	NO	NO
	FILE	screen,graph single case	graph,file	file <= 21 wnd spds 6 stab classes	file single case	graph,file single case
utes	10-30 minutes	inst-ave expose contin-not spec	user input <= simul period > puff interval	not specified	steady state	not specified but PG curves & steady wind
ied	5 Km	not specified	not specified	not specified	not specified	
. as of te, dist on	peak conc. as function of sigma-theta, delta-T & dist - based on 110 tests	79 chemicals heavy & simple gas submodels	EPA model	no user instr no doc ref. Hoot & Meroney(1973) wnd tunnl tst	stdy state ver no user instr puff vers under development	40 chemicals

(The reverse of this page is blank.)

information on 11 models currently on our microcomputers:

> AFTOX*, AVACTI II, CHARM*, D2PC, MADICT, OB/DG*, OME, PUFF/INPUFF, RVD, SLAB*, and SPILLS.

The asterisks indicate models that were developed either wholly or partially under the support of the U.S. Air Force.

The CHARM model is the most comprehensive of those on the list, but is a proprietary model which, because of partial support by the AFESC, is given to U.S. Air Force - affiliated users. The other models all have certain limitations that make them less comprehensive (or desirable) than several of the proprietary models in Table 5 (such as SAFER, HASTE, and MIDAS or CARE). The general characteristics and limitations of the models are briefly listed below.

AFTOX. This model was developed by AFGL (Reference 2) as a replacement for the OB/DG model. The AFTOX model, described in Section II-A, is a Gaussian model that does not treat dense gases. It will handle instantaneous or continuous evaporative emissions from spills, but will not handle jet releases. It has been extensively compared with field tests of continuous, neutrally buoyant emissions from the Prairie Grass, Green Glow, Dry Gulch, and Ocean Breeze sites.

AVACTA II. This is a puff model developed by Zannetti (Reference 53) that will account for variable wind fields in complex terrain. It is suitable for instantaneous, transient, or continuous releases. However, it cannot be used for dense gases.

CHARM. As stated above, the CHARM model (Reference 35) is the most comprehensive (by far) of the group of 11 models, but is not fully available.

D2PC. This is a U.S. Army model used for calculating the dispersion of specific munitions (Reference 9). The physical components of the model are excellent, but the user cannot get into the code to make use of these components. The user-friendly system is highly simplified and deals only with special munitions.

MADICT. This model (Reference 54) is very similar to AVACTA-II in that it is a puff model that accounts for variable wind fields in complex terrain. Dense gases are not treated.

OB/DG. The OB/DG model is a one-line empirical correlation developed by Nou (Reference 27) for application to Air Force sites where liquid propellants are stored. It is good only for the range of conditions used in its derivation:

Daytime stabilities
Ground-level continuous neutrally-buoyant source
Downwind distances less than about 5 km.

OME. This is a highly-simplified screening model used by the Ontario Ministry of the Environment (Reference 55) for application to a wide variety of hazardous gas releases. It can be used for dense gas jets or spills, but has been shown to have questionable performance when compared with field data.

PUFF/INPUFF. This EPA model (Reference 56) applies to puff transport and dispersion of non dense gases. It is similar to AVACTA-II and MADICT, but does not handle spatially variable wind fields in complex terrain.

RVD. The Relief Value Discharge (RVD) model of the EPA applies only to elevated continuous releases of dense gases, and is based on the Hoot-Meroney-Peterka model. The primary output is the concentration at the point the plume touches the ground. It includes an arbitrary conservative factor of five.

SLAB. The SLAB model is being modified by Ermak et al. (Reference 13) under support of the API and the AFESC so that it can operate on a PC and so that it can handle transient releases. It employs an evaporative source term and can account for dense gas slumping. The modifications are not yet complete.

SPILLS. The SPILLS model (Reference 56) does what it says and models evaporation from hazardous spills. However, it employs a simple Gaussian dispersion algorithm to calculate downwind dispersion that does not include the effects of dense gas slumping. It was used as a basis for the development of the AFTOX model.

Note: Raj et al. (Reference 38) are developing the ADAM model for AFGL that will account for two phase jet releases and dense gas slumping. It will be released in 1988.

D. REVIEW OF DATA SETS

Many field experiments have been conducted for the purpose of evaluating dispersion models. Draxler (Reference 58) reviews those carried out with positively or neutrally buoyant sources. Table 7 summarizes several of these experiments that utilized neutrally buoyant tracers. Hanna and Drivas (Reference 1) review those carried out with negatively buoyant sources, and Table 8 contains a summary from that document. The following subsections provide a review of specific data sets used later for evaluation of models or planned for model evaluation in Phase II of this project. The first section is a review of tests that used negatively buoyant tracers; the second section reviews neutrally buoyant tests. More detailed descriptions of the Prairie Grass, Green Glow, Dry Gulch, Ocean Breeze, and Thorney Island experiments are given in Appendix A.

1. Negatively-Buoyant Tracer Tests

a. ESSO/API, Gaz de France, DGA Netherlands

These tests, conducted in the early 1970s were of limited scope and applicability. All were on a much smaller scale than more recent experiments (Reference 59). ESSO/API and the Gaz de France used LNG released onto sea and soil surfaces, respectively. The DGA Netherlands tests used Freon-12 as the tracer and released it onto sand. The data compiled from these tests were not nearly as comprehensive as those from more recent tests. As these data are of limited use in model development or evaluation, they have not been compiled and put into the SRC database.

b. Porton Down

Forty-two releases of Freon-12, approximately 40 cubic meters in volume, were conducted at Porton Down, England, in 1978. The tests encompassed a range of initial gas cloud densities, wind speeds, and varying

TABLE 7. EXAMPLES OF NEUTRAL TRACER FIELD EXPERIMENTS.

Test and Reference	Tracer	No. of Tests	Release Type	Release Height	Averaging Time	Source Strength
Prairie Grass Barad 1958	SO_2	70	C	Ground	10 min	40.2-104.1 g/sec
Green Glow Barad & Fuquay 1962	FP	26	C	Ground	30 min	.85-7.04 kg/30 min
Ocean Breeze Haugen & Fuquay 1963	FP	76	C	Ground	30 min	.55-3.31 kg/30 min
Dry Gulch Haugen & Fuquay 1963	FP	109	C	Ground	30 min	1.1-3.3 kg/30 min
Sand Storm Taylor 1965	Be	43	QI	Ground	------	Variable
Cabauw Agterberg et al. 1983	SF_6	30	C	Elevated	30 min	1.06-4.61 g/sec

Symbol Legend

SO_2	Sulfur Dioxide
FP	Fluorescent Particles
Be	Beryllium
SF_6	Sulfur Hexafluoride
C	Continuous
QI	Quasi-Instantaneous

TABLE 8. EXAMPLES OF DENSE GAS FIELD EXPERIMENTS (REFERENCE 1).

Site and Reference	Material	No. of Tests	Q_1 (m^3 liquid)	Q ($m^3 min^{-1}$ liq)	Surface
DGA Netherlands van Ulden 1974	Freon 12	2	1000 kg	-	sand
Gaz de France Humbert and Montet 1972	LNG	40	-	0.16	soil
ESSO/API Feldbaner et al. 1972	LNG	17	.09-10.2	-	sea
HSE Porton Picknett 1982	Freon	35	40m^3 gas	-	grass-land
Maplin Sands Puttock et al. 1982	LNG, propane	34	27	1-5	sand, sea
China Lake Burro Koopman et al. 1982	LNG	8	40	12-18	pond
China Lake Coyote Goldwire et al. 1983	LNG	15	3-28	6-19	pond
Desert Tortoise Koopman et al. 1984	NH_3	4	60 over 7.5 min.		sand
Eagle	N_2O_4	6	4.2 over 3 min.		sand
Thorney Island Puttock and Colen-brander 1985	Freon	28	2000 instantaneous		airport

degrees of surface roughness. Concentration measurements were made by eight monitors located on two masts 25 meters downwind and 5 degrees to either side from the expected cloud centerline. Wind data were measured by 10 sets of anemometers and nine wind vanes, all of which were designed at the Chemical Defense Establishment at Porton. Temperature and relative humidity were also measured using two aspirated psychrometers located 0.5 and 4 meters above ground level. Estimates of the Pasquill stability class were made and the roughness length was determined.

c. Maplin Sands

Thirty-four spills of refrigerated liquid propane and liquified natural gas were conducted in an area of tidal sands on the north side of the Thames River estuary in England during 1980 (Reference 40). Shell Research Limited performed the tests under conditions of offshore winds for safety purposes. The release point was 350 meters offshore; if possible, spills were made at high tide. Gas sensors were located on 71 floating pontoons, usually at heights of 0.5, 1.4, and 2.4 meters above the sea surface. Other instrumentation included thermocouples and sonic anemometers. Two special pontoons provided vertical profiles of wind speed and temperature up to 10 meters; additionally, wind direction, relative humidity, solar insolation, water temperature, and wave height were measured. Detailed information on several of these spills are included in the computer data base set up at Sigma Research Corporation.

d. Burro

The Burro series of nine tests were held at China Lake, California, in 1980. Forty cubic meter spills of liquified natural gas of varying duration were conducted under the sponsorship of the U.S. Department of Energy and the Gas Research Institute (Reference 40). Gas sensors and thermocouples were mounted at three heights on 30 stations located on four arcs 57, 140, 400, and 800 meters from the release point. Five of the 30 stations had three anemometers mounted as well. Wind data were also collected from an array of 20 stations that each had an anemometer mounted at 2 meters above ground level. The water basin in which the LNG was spilled was 58 meters in diameter and 1 meter deep. The terrain was relatively flat, sloping

slightly downward from left to right as viewed looking downwind from the release point. Detailed information on four of these spills are included in the SRC database.

e. Coyote

The Coyote series of ten tests were conducted at China Lake, California in 1981 (Reference 40). Spill volumes of LNG were fairly small and both spill volume and duration varied over these tests, which were sponsored by the U.S. Department of Energy and the Gas Research Institute. Thirty stations instrumented with gas sensors and thermocouples were arrayed between 100 and 500 meters downwind from the source. Wind data was collected from an array of 20 stations, each with an anemometer mounted at 2 meters above ground level. The water basin in which the LNG was spilled was 58 meters in diameter and 1 meter deep. The terrain at the test site was relatively flat and sloped slightly from left to right as viewed looking downwind from the source. Detailed information on three of these spills are included in the SRC data base.

f. Desert Tortoise

Four releases of NH_3 were conducted at Frenchman Flat, Nevada, in 1983 by the Lawrence Livermore National Laboratory (Reference 40). All tests were done under conditions of constant pressure and flat terrain with both stable and neutral atmospheric conditions. Vapor cloud concentration and temperature were measured on arcs located 100, 800, and 1400 meters downwind. Wind data were collected from cup and vane anemometers located both upwind and downwind of the source at 2 meters above ground level. Detailed information on two of these spills are included in the SRC data base.

g. Eagle

Six releases of nitrogen tetroxide (N_2O_4) were conducted at the Department of Energy Nevada test site in 1983 by the Lawrence Livermore National Laboratory for the U.S. Air Force (Reference 15). Vapor cloud concentration and temperature data were measured on 5, 10 meter towers located 785 meters downwind of the source. The sensors were located 1, 3.5, and 8.5

meters above ground level. Ten second average values of wind speed, direction, and standard deviation of direction were determined at nine locations from two-axis, cup-and-vane anemometers sited at 2 meters above ground level. Information on two of these spills are included in the SRC data base.

h. Thorney Island

Thorney Island, England, was the site of 16 unobstructed, large-scale releases of a heavy gas tracer during the summer of 1982 and 1983 (Reference 48). The experiment was conducted by the National Maritime Institute under contract to the United Kingdom Health and Safety Executive. Instantaneous releases of 2000 cubic meters of a gas mixture of Freon-12 and nitrogen were accomplished using an accordion-type container. Tracer concentrations were measured using gas sensors placed at heights of 0.4, 2.4, 4.4, and 6 meters above ground on 38 fixed and four mobile masts located in a 100-by 100-meter grid. Meteorological measurements included wind speed and direction, turbulence, temperature, pressure, relative humidity, and solar insolation. Pasquill-Gifford stability classes were determined.

The test site was an abandoned airfield complete with runways. The locale was flat and uniform over an area of 1 by 0.5 kilometers. The surface was grass interspersed with tarmac-runways. The roughness length was determined to be 1 centimeter. Wind data were measured at 10 meters above ground level. Information on five of these tests is included in the SRC data base.

2. Neutrally-Buoyant Tracer Tests

a. Project Prairie Grass

Project Prairie Grass, designed by Air Force Cambridge Research Center personnel, was held in north central Nebraska near O'Neil in the summer of 1956 (Reference 60). SO_2 was released continuously over 10-minute periods from ground level in the 70 trials that comprised the project. Dosage measurements were made on arcs located at 50, 100, 200, 400, and 800 meters downwind. About half of the trials were conducted during unstable daytime

conditions and the rest were held at night with temperature inversions present. Meteorological measurements included wind speed, direction, and fluctuations in direction from cup anemometers and airfoil type wind vanes. Micrometeorological data, rawinsonde data, and aircraft soundings were also taken.

The site was located on virtually flat land covered with natural prairie grasses (latitude 42° 29.6'N, longitude 98° 34.3'W). The roughness length determined for the site by some of the researchers was 0.6 centimeters. Dosages were measured at a height of 1.5 meters along the arcs using midget impingers. The meteorological data were given in 10-minute averages. Information on all of these tests are included in the SRC data base.

b. Project Green Glow

Project Green Glow, a joint program designed by the Hanford Laboratories of General Electric and the Air Force Cambridge Research Laboratories, was held at the Hanford reservation in south central Washington in 1959 (Reference 61). Fluorescent particles (which gave off a green glow under ultraviolet light) were released continuously over 30-minute periods, using aerosol fog generators in the 26 trials that made up the experiment. Tracer dosages were measured using membrane filters at arcs 200, 800, 1600, 3200, 12800, and 25600 meters downwind. All of the trials were held at night under stable conditions. Meteorological measurements included wind speed and direction, temperature, and dew-point temperature from a 410-foot tower and a 78-foot mast and wind data only from 18 Hanford remote stations. Rawinsonde data were also taken.

The site was surrounded by elevated terrain and drainage flows were common. The surface vegetation consisted of desert grasses interspersed with sagebrush 1 to 2 meters in height. No roughness length values were given in the project report; from site descriptions, it is estimated to be from 1 to 3 centimeters. Tracer dosages were measured using a Rankin counter and a multiplier phototube. Special arraying procedures were used on samples that had been contaminated by blowing dust. Information on all of these tests are available in the SRC data base.

c. Project Ocean Breeze

Project Ocean Breeze was conducted at Cape Canaveral, Florida, by Air Force and General Electric personnel during 1961 and 1962 (Reference 29). Fluorescent particles were released continuously from ground level for 30-minute periods using aerosol fog generators in the 76 trials comprising the project. Tracer dosages were measured at arcs located 0.75, 1.5, and 3 miles downwind at a height of 15 feet above ground level. Many of the trials were run under sea breeze conditions. Meteorological measurements included wind speed and direction using Belfort devices sited 12 feet above ground and temperature profiles from captive wiresonde instrumented balloons. Standard synoptic and rawinsonde data were also provided.

The test site was on the missile range located on the east-coastal Florida coast. The locale was characterized by 10-20 feet tall rolling sand dunes. In addition, much of the diffusion course was covered with brushwood and palmetto growth. No roughness length value was given in the project report. Tracer dosage samples were assayed using a Rankin counter and a multiplier phototube (as in Project Green Glow). Information on all of these tests are available in the SRC data base.

d. Project Dry Gulch

Project Dry Gulch was conducted at Vandenberg Air Force Base, California, by Air Force and General Electric personnel during 1961 and 1962 (Reference 29). Fluorescent particles were released continuously from ground level over 30-minute periods using aerosol fog generators in the 109 trials that comprised the project. Tracer dosages were measured with membrane filters on arcs located 2301 and 5665 meters downwind on diffusion course B and 853, 1500, and 4715 meters downwind on Course D. Many of the trials were run under sea breeze conditions. Meteorological measurements included wind data from Belfort devices placed 12 feet above ground level and temperature data from wiresonde devices. Rawinsonde data, including many special launches, were also provided.

The test site was located on Burton Mesa. The terrain was quite complex; no stretch of the imagination would deem it flat. No roughness

length value was given in the project report, but it would undoubtedly be high due to the terrain. Tracer dosages were assayed using Rankin counters and multiplier phototubes. Information on all of these tests are available in the SRC database.

e. Project Sand Storm

Forty three, quasi-instantaneous releases of rocket solid propellant mixed with metallic beryllium were carried out at Edwards Air Force Base, California (Reference 62). All tests were carried out under unstable atmospheric conditions. Rocket motor firing times ranged from 2 to 8 seconds. Initial puff diameters of 15 to 45 meters were observed visually. Membrane filters were used to collect the tracer along arcs from 100 to 2400 meters downwind. After the first 14 trials, the samplers were consolidated onto six arcs located from 200 to 2400 meters away from the source.

Meteorological data were collected on an instrumented, 60 meter tower which was located 60 meters upwind of the release point. Wind speed and direction as well as temperature data were collected at multiple levels. The standard deviation of the wind direction, σ_θ, was also measured.

f. Cabauw, Netherlands

A series of dispersion tests were carried out at Cabauw in the Netherlands between 1977 and 1978 under the auspices of the Royal Netherlands Meteorological Institute and the KEMA Laboratories. The complete data sets of 15 trials are to be found in Reference 63. These tests consisted of elevated releases (80 or 200 meters) of SF_6 from the instrumented 213-m mast. Concentration measurements were made on an arc 4 kilometers downwind. Meteorological measurements included wind speed and direction, temperature, and turbulence from the instruments on the mast (Reference 64), radiosonde, sonar, synoptic observations, and radiation measurements.

The mast is located in the flat, central portions of The Netherlands (longitude 51°58′ N and latitude 4°56′E) between Schoonhoven and Lopik. The surrounding terrain, while flat for a radius of about 20 kilometers, is dotted with small villages, lines of tress, river dikes (the

river Lek is about 1 km away at its closest point), and meadows. Owing to the variety of surface covering around the mast, the surface roughness legnth, z_o, varies with wind direction and time of year from 6 to 25 cm.

Each of the 15 trials were composed of two consecutive, 30 minute tests; the corresponding mast data is presented as 30 minute averages. Data analysis and comparisons with other tests can be found in References 65 and 66. All of these data are contained in the Sigma Research data base.

SECTION III

COMPONENTS OF MODEL UNCERTAINTY

A. OVERVIEW OF THREE COMPONENTS OF MODEL UNCERTAINTY

Much work has been done on model uncertainty by researchers in other areas, including economics, ecology, and health sciences. These persons must also deal with widely scattered data, incomplete input data, non-Gaussian distributions, and wide confidence bounds. The following discussion summarizes the general approach to model uncertainty that has been adopted for this research, which closely parallels recommendations by Hanna (Reference 38).

If major decisions are to be made regarding the choice of an appropriate model of hazardous gases, evacuation plans, and risk assessments, it is important to have the best possible information on our confidence in the models that are used and the data that are being collected. It may even be possible to build the confidence intervals (uncertainty) into the decision-making process. There are three components of total error or uncertainty in models for source emissions, transport, and dispersion of hazardous gases.

Errors caused by model physics assumptions
Random variability (turbulence)
Errors generated by input data errors

These components have not yet been studied in any comprehensive way. Our general philosophy on model evaluation and uncertainties is shown in Figure 10, where the three components of uncertainty are plotted as a function of the number of parameters in the model. The total uncertainty can be large for models with a large number of parameters, due to the combined effect of data input errors. It is desirable to construct a model such that the total model uncertainty on the figure is at its lowest point. It is seen that a complex model is not always the best model in a given application. In many instances

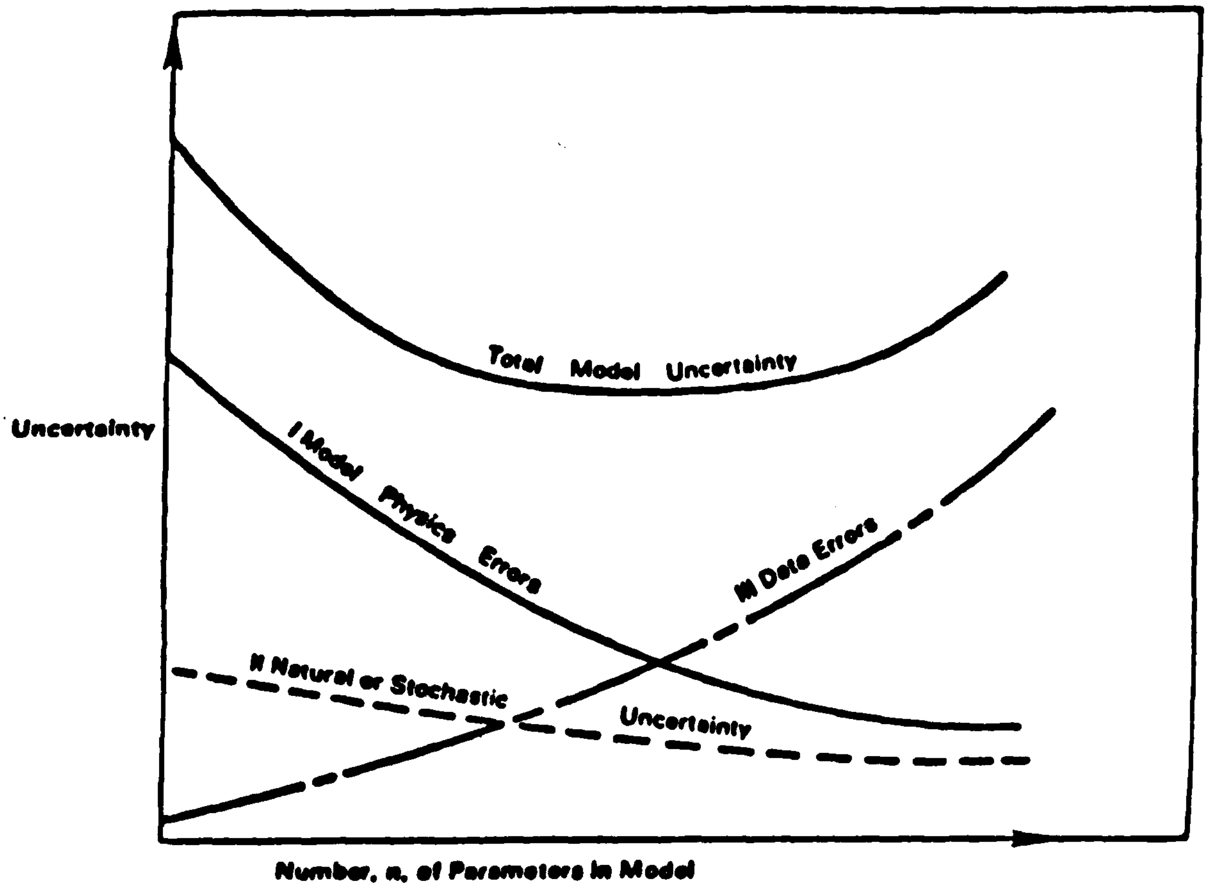


Figure 10. Illustration of Variation of Model Uncertainty Components with Number of Parameters in Model.

a Gaussian dispersion model or a simple scaling relation (e.g. $C \propto Q/ux$) provides the lowest uncertainty.

The uncertainties can be quantified by defining total uncertainty as the mean square residual, $\overline{(C_p-C_o)^2}$, and assuming that C_p and C_o are given by:

$$C_o = C_{oa} + C_o' + \Delta C_o \tag{5}$$

$$C_p = \bar{C}_p + C_p' + \Delta C_p \tag{6}$$

where C_p is the predicted parameter (for example, concentration) and C_o is the observed parameter. Other definitions are given below:

C_{oa} is the ensemble average that would apply to these external conditions

C_o' is the stochastic (random) fluctuation about the ensemble average

ΔC_o is the data error in the observation C_o

$\bar{C}_p$ is the predicted ensemble average

C_p' is the predicted stochastic (random) fluctuation about the ensemble average

ΔC_p is the error due to data input errors

No current hazardous gas models predict the stochastic fluctuation C_p'. If it is assumed that there is no correlation among the components, then the total uncertainty or the sum of the three components shown in Figure 10, is given by subtracting Equation (5) from Equation (6), squaring the difference, and averaging (overbar):

$$\overline{(C_p - C_o)^2} = \overline{(\bar{C}_p - C_{oa})^2} + \overline{\sigma^2_{C_o}} + \overline{\Delta C_o^{\,2}} + \overline{\Delta C_p^{\,2}} \qquad (7)$$

Total Model Uncertainty	Physics Error I	Stochastic Uncertainty II	Data Errors III

Observations have shown that the stochastic uncertainty component II is roughly equal to the square of the mean (i.e. $\overline{\sigma^2_{C_o}}/\bar{C}_o^2 \sim 1$) for small averaging times (for example, a few seconds) and can be predicted by some models (for example, Reference 67). The data errors component III is also believed to be of the same order as the square of the mean. The model physics error component I can be estimated by solving Equation (7) for that component, given the total model uncertainty and components II and III. Of course, this procedure is highly uncertain if the components II and III are approximately equal to the total model uncertainty.

B. DATA INPUT UNCERTAINTIES

The data errors component III can be estimated based on studies of instrument errors in the field. It is stressed that some QC (quality control) procedures, such as checking the voltage output of an anemometer at a given rotation rate, do not tell much about actual instrument error in the field. These actual errors are best determined through use of co-located instruments and comparision with high quality "base-line" instruments. In most field experiments, this option is not practical.

1. Uncertainty in Tracer Gas Observations

Information on data errors in tracer gas observations was obtained using co-located instruments as part of the study of dispersion from tall stack plumes sponsored by the Electric Power Research Institute (EPRI). Table 9 gives some figures on uncertainties for source emissions and tracer gas observation instruments (Reference 68).

TABLE 9. UNCERTAINTIES IN TRACER GAS OBSERVATIONS AT EPRI KINCAID SITE (FROM REFERENCE 67)

Parameter	Standard Deviation or Other Limit
Source SF_6	3%
Monitored SF_6	6% for $C > 100$ ppt, 6 ppt for $C < 100$ ppt
Wind direction to max. observed C	20°

The wind direction error is obtained by calculating the difference between the observed wind direction and the direction from the source to the receptor with the maximum observed concentration. The site consists of flat farmland with some manmade lakes interspersed in the area, and the instruments were all research grade monitors operated by technicians. Thus these data errors represent the minimum of what would be found in a monitoring network in the neighborhood of an industrial site. In terms of Equation 7, the minimum value of $\left[\bar{\sigma}_{C_o}^{2} / \bar{C}_o^{2}\right]^{1/2}$ is expected to be about 0.06. It is important to note that this relative error grows quickly at concentrations near the threshold of the instrument.

2. Uncertainty in Meteorological Data

Typical uncertainties associated with the measurement of meteorological data have been addressed in several studies. The results from five studies are reviewed here.

The earliest of this group of studies is the Prairie Grass project. Although several different organizations were involved in making meteorological measurements, the project report (Reference 60) cites accuracy assessments for only the slow response MIT measurements taken at an elevation of 2 meters. The second reference reviewed is the report from the workshop on on-site meteorological measurements (Reference 69). Although not concerned with any one specific field program, the attendees reported on their collective experience in making meteorological measurements. The third reference is a study that was specifically designed to compare measurements obtained from 5 types of mechanical wind sensors, and a sonic anemometer (Reference 70). The test instruments were mounted on separate 10 meter tall

masts, set approximately 5 meters apart (across the flow). The fourth reference assesses measurements made in support of the EPRI plume model validation study at the Kincaid and Bull Run sites (Reference 68), while the fifth includes a report on measurements made at Dugway Proving Grounds (Reference 71). The latter report is particularly interesting in that comparisons are made between "identical" wind instruments installed 500 meters apart.

Table 10 presents information obtained from each of the references. All of the references noted that the mechanical wind sensors are not as reliable for wind speeds of less than approximately 2 m/s because of starting thresholds and response times. As stated in the workshop report (Reference 69), typical thresholds attainable with cup and vane instruments are on the order of half a meter per second. The workshop report also brought up the subject of sampling rate. For averaging times of 60 minutes, 60 or more samples will estimate the mean to within 5 to 10 percent, and 360 or more samples will estimate the standard deviation to within 5 to 10 percent. Because the sampling density and accuracy are related to the time-scales of the measured variable, similar statements for shorter averaging times such as 10 to 20 minutes are not applicable. Most data collection systems used in these studies sample the speeds and directions at least once per second.

The BAO study results (Reference 70) are predicated on the belief that the sonic anemometer provides the best attainable measurements of wind speed and direction, so that measurements from the mechanical systems are only compared with those from the sonic anemometer. The scatter found in wind direction measurements (4.5°), relative to the sonics, is surprisingly large. This caused one of the authors of the study to look at the comparability of the mechanical systems among themselves. His analysis is not complete, but in studying one 20-minute period, he finds that the vanes indicate directions to within 1°, bivanes to within 1.5°, and cups indicate speeds to within .1 m/s, when compared with the prop-vane used in the study. No justification has been presented for discounting the discrepancy between the sonics and the mechanical systems.

With regard to the σ_{θ} data obtained from the BAO report, we note that the bias in σ_{θ} (vane) as a function of σ_{θ} (sonic) departs from near-zero rapidly

TABLE 10. TYPICAL UNCERTAINTIES IN METEOROLOGICAL MEASUREMENTS

	u	θ	σ_θ	σ_ϕ	ΔT[1]	Ave. Tikme
Prairie Grass (u > 2 m/s)	2-5%	2^o-5^o	10%	-	-	10 min.
Workshop	.2 m/s+5%	<3^o	5-10%	-	<0.1°C	60 min.
BAO[2]	.3 m/s	4.5^o	3^o	1.7^o	-	20 min.
BAO[3]	-	-	10%	-	-	20 min.
EPRI	.1 m/s	1.5^o	13%	20%	-	60 min.
EPRI (tower shadow)	10%	10^o	-	-	-	60 min.
DPG (mfg. specs).	.1 m/s+1%	3^o	1.2^o	-	.4°C	-
DPG (u > 5 m/s)	6%	-	-	30%	-	10 min.
DPG (u < 2 m/s)	25%	-	-	48%	-	10 min.
DPG (tower wake)	9% (u > 5 m/s)		-	-	-	10 min.
	<4% (u < 2 m/s)		-	-	-	10 min.

1. The temperature difference (ΔT) uncertainty applies to any height interval, since the instrument measures only a difference without regard to the location of the two points.

2. Mean bias removed (cups, props, vanes, bi-vanes relative to sonics).

3. Bias as a function of indicated σ_θ removed (vanes relative to sonics).

for σ_θ greater than 30°, and the standard deviation in the difference between the mechanical systems and the sonic grows at the same rate. To obtain a better estimate of the uncertainty in σ_θ measurements with a correction for the bias, valid throughout the range, we use a very simple representation of the bias curve for σ_θ greater than 30° to adjust the root-mean-square error that was reported. With the correction, the root-mean square error is approximately $1^\circ + \sigma_\theta/12$ for σ_θ greater than 30°. For σ_θ less than 30°, the root-mean-square error is approximately $\sigma_\theta/9$. Hence, the uncertainty in σ_θ(vane) relative to the measurements from the sonic anemometer is about 10%, as indicated in the table.

Both the EPRI (Reference 68) and DPG (Reference 71) reports include a measure of the effect of a support tower on wind measurements. Although this is an important consideration for general data accquisition requirements, it is of limited importance for short-term experiments in which the field study can be conducted only for periods in which the instruments are properly exposed. When exposure is a problem, both studies indicate that the effect on wind speed can be as large as 10 percent (indicated speeds are lower by about 10 percent).

The DPG report by White et al. (Reference 71) provides some information on uncertainties from lack of representative sampling. The word 'representative' refers to whether the measurement at a given location would agree with a concurrent measurement in the same general area. In looking at the effect of tower shadows or wake zones, they also present statistics for periods when the instrumentation on neither tower is in the shadow. Under these conditions, differences between "identical" instruments located 500 meters apart are more likely the result of differences in the flow, rather than differences in response or calibration. The data presented were obtained from cup anemometers, and bivanes mounted 16 meters above the ground. As indicated in Table 10, uncertainty in wind speed is about 6 percent for speeds greater than 5 m/s, and it rises to 25 percent for wind speeds less than 2 m/s. The other variable reported is σ_ϕ, the standard deviation of the vertical elevation angle. For speeds greater than 5 m/s, the uncertainty is found to be 30 percent. For speeds less than 2 m/s, the uncertainty grows to 48 percent. Although the data are not reported, the same type of information on θ and σ_θ should be available, and should be analyzed.

In general, the degree of uncertainty in wind measurements reported in these five documents is consistent. Without considering representativeness issues, wind speed measurements are generally within 5 percent of the "true" value, and wind directions are within 3° to 5°. The recent studies of lateral turbulence (σ_{θ}) also tend to confirm expectations based on experience (e.g., the workshop report and the Prairie Grass report), with an uncertainty in σ_{θ} of about 10 percent. But vertical turbulence (σ_{ϕ}) data appear to be more unreliable. It appears that σ_{ϕ} carries with it an uncertainty of about 20 percent, even when efforts are made to assure that the vane or prop is functioning properly.

When representativeness is considered, the uncertainty grows appreciably. Problems of exposure, such as tower shadows or wakes, can easily double the uncertainty in wind speed and direction. Worse, measurements taken at one point can differ substantially from measurements made using identical instrumentation located just 500 meters away. This is particularly true of turbulence measurements, and certainly wind speed measurements made under light wind speeds (less than 2 m/s). Although not documented, differences in wind direction are expected to be equally sensitive to spatial variations under light wind speed conditions.

Alternate methods for the measurement of vertical turbulence (σ_{ϕ}) are frequently employed, given the difficulty of obtaining reliable data on σ_{ϕ} from vanes and props. Stability class is sometimes used as an alternate method, inferred from observations of wind speed (near the ground), and surrogates for the sensible heat flux. An estimate of the uncertainty in σ_{ϕ} that arises in the use of these methods can be obtained by assuming that the resolution in the resulting stability class or category is no better than one half a class. The Briggs (Reference 72) dispersion parameter curves for σ_{ϕ} in rural areas contain a leading coefficient for each class that is essentially a mean σ_{ϕ} (in radians) for the category. If a linear trend is computed for these coefficients, from class B to class F, its slope is approximately .02 radians/class. Therefore, an uncertainty of half a class produces an uncertainty of approximately .01 radians, or 20 percent of the mean σ_{ϕ} for the entire range (class B to class F). Therefore, an uncertainty of about 20 percent would be associated with the use of surrogate methods (via the stability class) for σ_{ϕ}, if the variability in σ_{ϕ} within each class were

ignored. But Luna and Church (Reference 73), among others, show that the scatter in observed values of σ_ϕ associated with each stability class is so great, that any measured σ_ϕ could belong to any one of the stability classes selected on the basis of the surrogate methods.

3. Air Force Meteorological Data Errors

As part of this project, Capt. L. Key (AFESC) contacted several U.S. Air Force personnel in an attempt to determine QA/QC procedures and expected meteorological data errors. Capt. M. Davenport's memo of 27 October 1987 contains the results of this survey, which is briefly summarized below:

Average Air Force Base:

Wind Direction Errors:	$\pm 2^\circ$
Wind Speed Errors:	$\pm$ 1.5 KT
Hygrothermograph Errors:	$\pm 2^\circ$F (temp.)
	$\pm 1.5^\circ$F (dew point)

Calibration Interval - approximately 1 month.
Quality Control (i.e., double checking of data) - none.
Reporting procedures:
Wind direction is rounded to nearest 10° and 1 KT.
Actual reading is based on a 1-minute averaging period (Airways) or 10-minute averaging period (METAR)

Mesowind Networks at Vandenberg, Edwards, and Patrick:

Edwards - Maintenance of 19 - tower net unsatisfactory

Patrick - WS threshold - 2 m/s
WS errors $\pm$ .15 mi/hr
WD errors $\pm 3^\circ$
Temperature Some $\pm .35^\circ$C; Some $\pm 3^\circ$C

Calibration Interval - one month
Quality Control - Flagging software

Vandenberg -	WS Error	± 1 KT
	WD Error	± 2°
	Temp. Error	± 1°F

No routine calibration - only electronics checks.

C. STOCHASTIC OR TURBULENT UNCERTAINTIES.

Discussions of concentration fluctuations and the influences of averaging and sampling times are either nonexistent or very minimal in the research reviewed to this point. Some models grossly parameterize this effect (term II in Equation 7) by assuming that the ratio of the peak (fluctuating) concentration to the model predicted mean concentration is about two. Chatwin (Reference 74) pointed out that in many cases involving accidental releases of hazardous gases, the maximum short term (~1 sec) concentration is the most important variable to predict. Lung damage from H_2S can occur with one breath if the concentration is sufficiently high, and an explosion of gas from an LNG accident can occur if a spark is struck in a small volume of gas at the flammability limit. According to Chatwin the mean concentration predicted by the model can be irrelevant in these cases, since the probability distribution function (pdf) of concentration fluctuations in the atmosphere is characterized by a standard deviation at least as large as the mean. The relative magnitude of short-term concentration fluctuations ($\sigma_c/\overline{C}$) is the same order as the relative magnitude of short-term velocity fluctuations ($\sigma_u/\overline{U}$) in the atmosphere. The parameters σ_c and σ_u are the standard deviations of turbulent fluctuations in concentration and wind speed, respectively. It is assumed that averaging times are about 1 second and sampling times are about 10 minutes. Thus it is important to predict the upper end of the pdf for the H_2S and LNG incidents described above. Since Chatwin's article was published, a few other researchers have studied this problem, although a comprehensive operational model has not been derived.

Predictions of models such as AFTOX or DEGADIS can be thought of as ensemble means for certain averaging times. An ensemble mean is defined as the mean over an infinite number of realizations of a given experiment. The averaging time is usually implicit in the data used by the model and in its

formulations for treating the input data - for example, if hourly averaged wind and turbulence observations are used, then the predictions represent a 1-hour average. If the Pasquill-Gifford-Turner dispersion curves are used, then the predictions represent a 10 minute average, since data from 10 minute periods were used to derive the curves. In the case of instantaneous (puff) models, the predictions represent an ensemble mean only to the extent that a large enough set of experiments (20 or more) was used to derive the model. These experiments should be conducted under the same external conditions (that is, wind speed, stability, source term). For example, if it were possible to run the Thorney Island experiments long enough that 100 independent time periods of 10-minute duration could be found which all satisfy the following conditions:

$$4.8 < u < 5.2 \text{ m/s}, \quad 65\% < RH < 70\%$$

$$10^\circ < Ta < 12^\circ C, \quad 10^\circ < T \text{ surface} < 12^\circ C$$

$$-2 < \text{net radiation flux} < 2 \text{ watts/m}^2$$

$$(\rho_p - \rho_a)/\rho_a = 2, \quad h=10m, \quad R=10m,$$

then the observed concentration field averaged over these 100 experiments would approach an ensemble average. It is obvious that it is difficult operationally and financially to generate ensemble averages from atmospheric field experiments.

Thus the results of a single experiment, or even three or four experiments conducted under similar external conditions will likely differ (perhaps by as much as an order of magnitude) from the ensemble mean predictions of the model. If this happens, it is not an indictment of the model but may be a manifestation of the inherent stochastic variability of the atmosphere.

Wind tunnel experiments can be used to study variability, since it is easier to insure repeatability of experiments and thus create a large ensemble of data. On the negative side, the wind tunnel cannot simulate larger scale eddies and other phenomena that contribute to variability in the atmosphere.

Furthermore, the laboratory Reynolds number is not high enough to permit the establishment of an inertial subrange like there is in the atmosphere. Meroney and Lohmeyer (Reference 75) conducted extensive studies of dense gas clouds released in a wind tunnel and calculated the stochastic or random concentration fluctuation intensity, $\sigma_c/\bar{C}$, for various source volumes, wind speeds and downwind distances. An example of these results is plotted in Figure 11, showing that the average $\sigma_c/\bar{C}$ is about 0.3 in this wind tunnel. The value of $\sigma_c/\bar{C}$ in individual experiments range from 0.1 to 0.7. In contrast, Hanna (Reference 67) reports observed values of $\sigma_c/\bar{C}$ of 1.5 on the plume centerline and $\sigma_c/\bar{C}$ of 5.0 on the plume edges for a smoke plume released in the atmospheric boundary layer and concentrations averaged over one second..

The probability distribution function (pdf) of concentration fluctuations in the atmosphere has been studied by several persons (References 67, 76, and 77), and all agree that the distribution is non-Gaussian and is skewed towards higher concentrations. For hazardous gas analysis, we are usually interested in the probability $P(C>C_L)$ that the concentration is higher than some limiting value, C_L:

$$P(C>C_L) = \int_{C_L}^{\infty} p(C)\, dC \tag{8}$$

It has been suggested by various persons that the probability distribution function, p(C), can be approximated by a log-normal, clipped normal, or Gamma function. The exponential function is a special case of the Gamma function, and is quite good for intermittent clouds or plumes. Another important factor is the intermittency, I, which is defined as the fraction of time that non-zero concentrations are observed at a monitor. For the exponential distribution, $\sigma_c/\bar{C}$ equals one if the intermittency is unity. In the general case, the pdf is given by the formula:

$$p(C) = (I^2/\bar{C})\exp(-IC/\bar{C})+(1-I)\delta(0) \tag{9}$$

where the Dirac delta function $\delta(0)$ equals 1.0 at C equal to 0 and equals 0.0 elsewhere. This can be substituted into equation (8) to give:

$$P(C>C_L) = I\ \exp(-IC/\bar{C}) \tag{10}$$

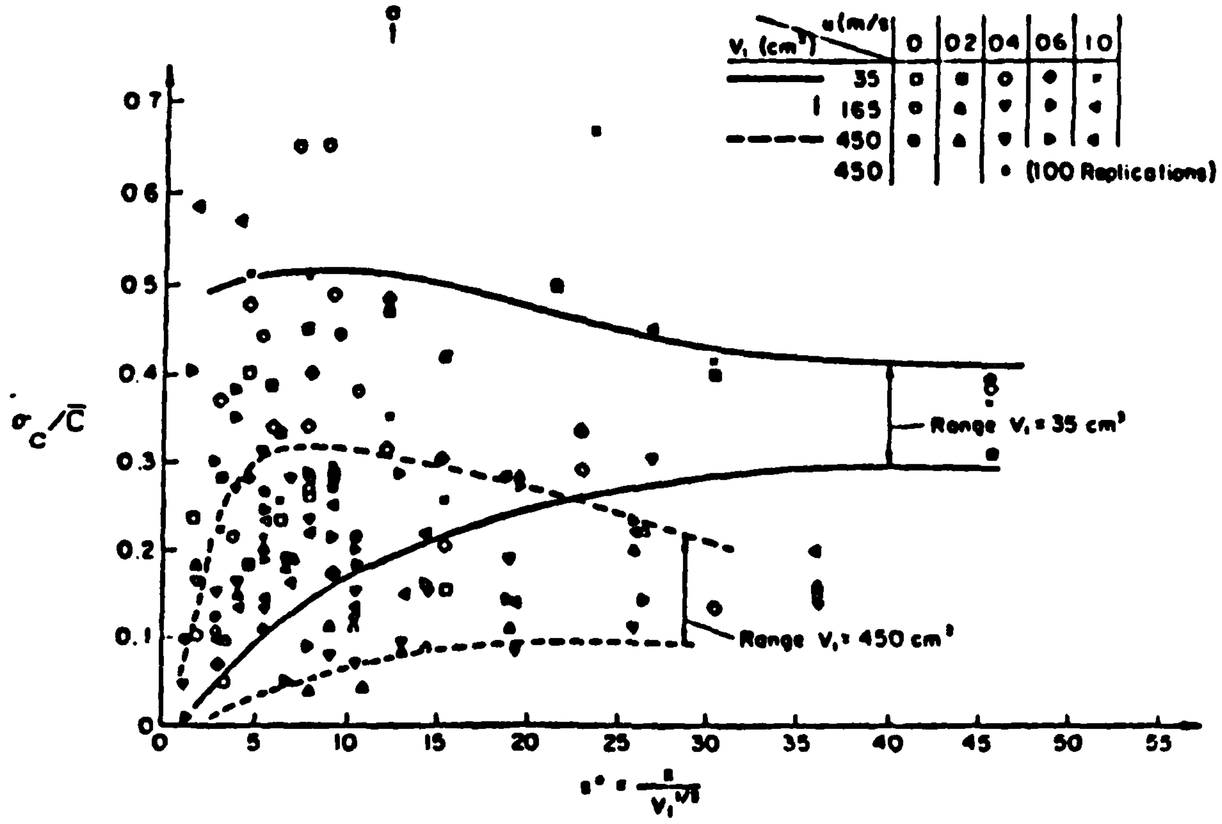


Figure 11. Concentration Variance Ratio, $\sigma_c/\overline{C}$, versus Downwind Distance, Observed by Meroney and Lohmeyer (Reference 75) in a Wind Tunnel. The source is an instantaneous dense gas cloud of initial volume V_1.

Thus, if the dispersion model predicts an ensemble mean, $\bar{C}$, of $0.1C_L$, where C_L is the threshold concentration for some health effect, and the intermittency I equals 0.5, then the probability that the instantaneous C will exceed C_L is 0.3 percent. If the ensemble mean prediction is $0.5C_L$, then this probability is 18 percent.

The formulas given above are for nearly-instantaneous averaging times. It is clear that the standard deviation of concentration fluctuations, σ_c, will decrease as averaging time T increases. If the integral time scale of the concentration fluctuations is T_I and their autocorrelogram is assumed to be exponential, then the following formulas apply:

$$R(t') = \overline{C'(t)C'(t+t')}/\sigma_c^2 = \exp(-t'/T_I) \tag{11}$$

Then

$$\sigma_c^2(T)/\sigma_c^2(0) = 2(T_I/T)(1-(T_I/T)(1-\exp(-T/T_I))) \tag{12}$$

where $\sigma_c^2(0)$ refers to the variance for instantaneous averaging time. If T_I has a typical value of 10 sec for the surface layer then the ratio of variances for an averaging time of T equal to 60 sec is 0.28. This estimate of T_I is based on observations of concentration fluctuations during smoke/obscurant experiments conducted by the U.S. Army (Reference 67). If the averaging time is one hour, the ratio $\sigma_c(3600s)/\sigma_c(0)$ is 0.075. It can be concluded that the fluctuation intensity $\sigma_c/\bar{C}$ for 1-hour averages in the atmosphere is about 0.1 even if the integral time scale is only a few seconds. For fluctuations dominated by lateral meandering, where T_I is more like 100 to 1000 seconds, the fluctuation intensity $\sigma_c/\bar{C}$ for one hour averages is approximately 0.5.

If it is assumed that the equations in the first part of this section produce predictions of ensemble mean concentrations, $\bar{C}$, then the probability of the concentration exceeding any threshold limit, C_L, can be estimated using equations (8) through (12) for any averaging time and integral time scale.

Equation (12) can be used to assess the effects of averaging over distances as well as time. Observed concentrations and health effects always involve some averaging distance. For example, if the integral distance scale

of the turbulence is 5 meters and the averaging distance is 1 meter, then the ratio $\sigma_c^2(1m)/\sigma_c^2(0)$ equals 0.94.

At the other end of the scale the sampling time or sampling volume can also influence observations. The sampling time, T_s, can be thought of as the total length of time that the instrument is turned on. The likelihood of more extreme concentrations being observed is increased if the sampling time increases (for example, several new "record" high and low temperatures are observed at any given weather station each year). The usual definition of any ensemble assumes that the sampling time is infinity. In practice this requirement is considerably relaxed, such that a set of 10 dense gas experiments conducted during similar external conditions is assumed to comprise an ensemble. Equation (12) can also be used to calculate the variance "missed" by an instrument because it is turned on for a finite sampling time T_s:

$$\sigma_c^2\ (0,\ T_s)/\sigma_c^2\ (0,\ \infty) = 1-2(T_I/T_s)(1-(T_I/T_s)((1-\exp(-T_s/T_I))) \qquad (13)$$

where the first variable inside the parentheses after σ_c^2 is the averaging time and the second variable is the sampling time. Any eddies with time scales much larger than T_I are not detected by the instrument. For example, if T_s is 10 times the integral scale T_I, then only 82 percent of the total possible variance is seen. If both the sampling time T_s and the averaging time T are finite (as they are in any experiment) then the fraction of the total possible variance can be calculated by multiplying Equations (12) and (13) together. An example is given in Figure 12 for the special case $T_s/T = 100$ (for example, averaging time could be one minute and sampling time could be 100 minutes). This function clearly defines a "window", with high and low frequency fluctuations filtered out by the finite sampling and averaging times.

D. MODEL PHYSICS ERRORS IN TYPICAL HAZARD RESPONSE MODELS

The most difficult component of the uncertainty to evaluate is the component due to model physics errors. This can be calculated as a resultant from Equation (7), but has a great potential for uncertainty itself because the stochastic and data input uncertainties are of the same order of magnitude as the total uncertainty. To determine whether the inclusion of a specific model physics component (along with the resultant increase in uncertainty that

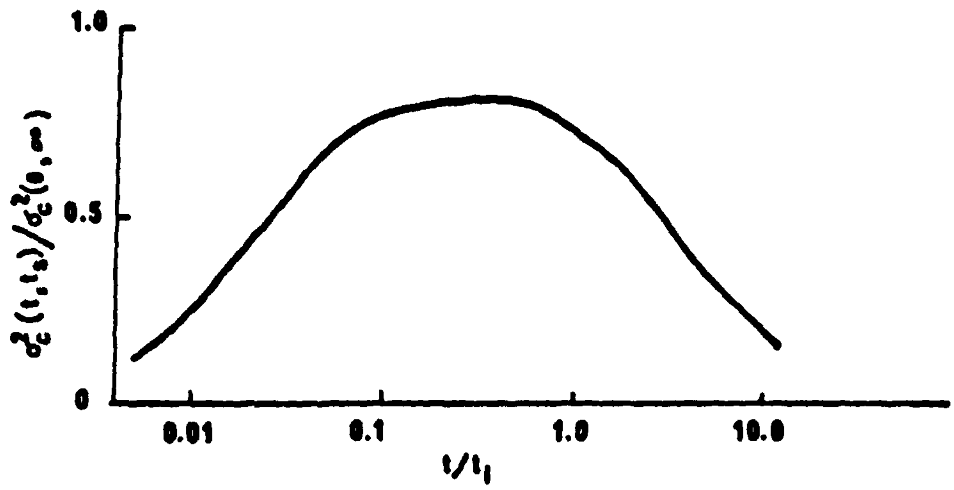


Figure 12. Fraction of Total Possible Concentration Fluctuation Variance from Equations (9) and (10), as a Function of Sampling Time T_s, Averaging Time T and Integral Time Scale T_I. It is Assumed that T_s/T_I = 100.

is associated with its input data requirements) reduces or increases the model errors, we use the model performance parameter C_o/C_p, where C_o is the observed concentration and C_p is the concentration predicted by some component or model. The variance of C_o/C_p is calculated and defined as var(Model i), where Model i could be defined in any manner. Rafferty and Dumbauld (Reference 78) also looked at the variance in the ratio C_o/C_p, from the perspective of assessing the variability in the source term among many smoke dispersion trials.

If Model 1 is given by the simple relation $C_p = 1$, then var(Model 1) equals the variance of the observed concentration C_o. Obviously it is desired that the variance of any other model be less than this amount. If the new model increases the variance, then it clearly has not demonstrated any skill. Other models can be quite simple or can be defined by standard models such as OB/DG or AFTOX:

Model 1: $C_p = 1.$
Model 2: $C_p = aQ$
Model 3: $C_p = bQ/u$
Model 4: $C_p = c/u$
Model 5: $C_p = dQ/(sigw*sigv*x^2*u)$
Model 6: $C_p = OB/DG$

where a, b, c, and d are constants.

We use our knowledge of dispersion to formulate the first five models. If the variance of, say, C_o/C_p for Model 3 is not less than the variance of C_o/C_p for Model 2, then this addition of the wind speed, u, is not important for the model predictions compared to the uncertainty it introduces (the uncertainty in measuring the representative value of the transport speed). An application of this procedure using the Prairie Grass dispersion observations is given below.

The data from the Prairie Grass experiments provide us with a set of tracer concentration and meteorological data for 53 trials. Data from other Prairie Grass trials are available, but these data are incomplete, and estimation of the stability class using the Golder (Reference 79) method was applied only to the set of 53 trials that we are analyzing. The stability

class is important in that wind fluctuations in the vertical were not measured, so these need to be inferred from the stability class index.

Models for C_p are evaluated first by taking the independent variables separately. For each model, the variance in the ratio C_o/C_p is computed, and scaled by the average ratio (squared) so that the variance associated with each of the models can be compared. That is, the scaled variance is given by:

$$\mathrm{svar}\ (C_o/C_p) = \frac{\mathrm{var}\ (C_o/C_p)}{(\overline{C_o/C_p})^2}$$

These scaled variances are reported in Table 11.

The first model, $C_p = 1$, provides us with the scaled variance in the observed concentrations, which equals 2.2 in this case. Each of the remaining models tests the importance of including emissions and plume-growth variables one at a time. Note that the standard deviation of vertical angle fluctuations, σ_ϕ, is inferred from the stability class by taking the leading coefficient of the curve for σ_z proposed by Briggs (Reference 72) for each stability class, and interpolating. Each variable except the emission rate, Q, and the square of the distance, x^2, reduces the scaled variance. This suggests that a model that uses either Q or x^2 may not perform well.

This conjecture is evaluated by formulating several models that combine the variables. These are listed in Table 12. The first combines the two variables that exhibited the greatest reduction in the scaled variance, wind speed, u, and distance from the source, x. Their product produces a relatively small scaled variance of 0.666. When the wind fluctuation variables σ_ϕ and σ_θ are introduced as well, the scaled variance rises to 0.753. This combination would suggest the following square root dependencies of σ_y and σ_z: $\sigma_y = \sigma_\theta\sqrt{x}$ and $\sigma_z = \sigma_\phi\sqrt{x}$. However, application of this model in the third line of the table shows no improvement over the simpler model. The next pair of models are similar to the previous pair, except distance is introduced as square (linear growth in σ_y and σ_z with distance). Now it is evident that even though the x^2 model performed poorly in Table 11, it performs very well when coupled with the turbulence variables, producing a scaled variance of 0.514. Thus, it appears that σ_y and σ_z are each

TABLE 11. COMPARISON OF SCALED VARIANCES OF C_o/C_p FOR SEVERAL SINGLE-VARIABLE MODELS OF C_p (PRAIRIE GRASS DATA).

Model for C_p	Scaled Variance
$C_p = 1$	2.200
$C_p = Q$	2.979
$C_p = 1/u$	1.652
$C_p = 1/\sigma_\theta$	1.875
$C_p = 1/\sigma_\phi$	1.860
$C_p = 1/x$	1.356
$C_p = 1/x^2$	4.674

* Scaled Variance = Variance $(C_o/C_p) \Big/ \overline{(C_o/C_p)}^2$

proportional to x. A third model for plume growth appears next. In this case, σ_y and σ_z are computed from the stability class index and the Briggs (Reference 72) curves, where interpolation between the curves is done for fractional stability class values. This model does not reduce the variance as much as the linear growth model. Finally, the results of the AFTOX model for unit emission rate are included, and this model is most successful in reducing the scaled variance (0.484). Apparently, the additional model physics incorporated in AFTOX relating to source-receptor geometry and plume growth rates is important in modeling these Prairie Grass data. In future research, confidence limits on these scaled variances should be calculated to determine whether the differences between, say, AFTOX and the $1/\sigma_\phi\sigma_\theta x^2u$ models are significant.

The last three models included in Table 12 contain the emission rate, and the scaled variances increase substantially as a result. Apparently, the effective emission rates for these trials are not well known, and including the emission rate in a model for these data actually degrades model performance. Use of a mean emission rate in AFTOX would produce better agreement with the observations than the use of the stated emission rate for each of the trials.

TABLE 12. COMPARISON OF SCALED VARIANCES OF C_o/C_p FOR SEVERAL MODELS OF C_p THAT INVOLVE COMBINATIONS OF MORE THAN ONE VARIABLE (PRAIRIE GRASS DATA)

Model for C_p	Scaled Variance
$C_p = 1/(xU)$	0.666
$C_p = 1/(xU\sigma_\theta\sigma_\phi)$	0.753
$C_p = 1/(x^2U)$	1.804
$C_p = 1/(x^2U\sigma_\theta\sigma_\phi)$	0.514
$C_p = 1/(U\sigma_y\sigma_z)$	0.597
C_p = AFTOX/Q	0.484
$C_p = Q/(x^2U\sigma_\theta\sigma_\phi)$	1.030
$C_p = Q/(U\sigma_y\sigma_z)$	1.103
C_p = AFTOX	0.933

* Scaled Variance = $\mathrm{Var}(C_o/C_p) \Big/ (\overline{C_o/C_p})^2$

SECTION IV

FRAMEWORK OF MODEL EVALUATION PROCEDURE

A. OVERVIEW OF MODEL EVALUATION APPROACH

A major part of this effort involves the development of a framework for the evaluation of currently available microcomputer-based hazard response models. This evaluation is intended to employ standard statistical procedures, and provides information only concerning the total model error. No information can be obtained on the components of the model error discussed in the previous section. Using the results of this statistical evaluation, it is possible to quantify the uncertainty for various scenarios. A framework for model evaluation has been developed that contains several options at certain key steps. These options have been tested in a preliminary way in Phase I, and may be further refined if Phase II is carried out. The steps in this framework have the following form:

1 Choice of models to be tested, listing of required input data, averaging and sampling times..

↓

2 Acquisition of data sets, generation of modeler's data base (i.e. appropriate input data for each period), estimation of input data uncertainty.

↓

3 Calculation of concentrations by each model.

↓

4 Calculation of model performance measures and confidence limits on these measures.

↓

5 Decision on whether the model predictions are significantly different from another model.

↓

6 Quantitative assessment of uncertainty of each model.

Sigma Research has been working on several model evaluation procedures that could be used as options in steps 4 and 5 of the framework (References 38 and 68). Other potential procedures have been suggested by Fox (Reference 80), the EPA (Reference 81), and Cox and Tikvart (Reference 82). Specific model performance measures are listed in Section IV.8. In general the purpose of this exercise is to answer two questions:

1. Are the model predictions significantly different from the observations?

2. Are the predictions of two models significantly different from each other?

The answers to these questions depend on prudent applications of statistical procedures for calculating confidence limits.

B. DESCRIPTION OF PERFORMANCE MEASURES

For many purposes a straightforward statistical analysis of model performance is necessary. In this section the recommendations of an EPA/AMS committee (Reference 80) are followed to the extent that a limited number of performance measures is employed. Accurate calculation of confidence limits is emphasized. Scientific review of the models is just as important as statistical evaluations.

1. Description of Statistical Analysis

A scheme for evaluating air quality models has been developed by Hanna and Heinold (Reference 83) which uses a modest set of performance measures and which emphasizes estimation of confidence limits on each performance measure. Two basic performance measures are used. The first is called the fractional bias, FB, and emphasizes the bias in the mean predicted concentrations (see Reference 82):

$$FB = (\overline{C_o} - \overline{C_p})/(0.5(\overline{C_o} + \overline{C_p})) \tag{14}$$

The second performance measure emphasizes the scatter in the entire data set and is defined as the normalized mean square error, NMSE (see Reference 83):

$$NMSE = \overline{(C_p - C_o)^2}/\overline{C}_o\overline{C}_p \tag{15}$$

The normalization by $\overline{C}_o\overline{C}_p$ assures that NMSE will not be biased towards models that overpredict or underpredict.

The data used for model testing should be independent of the data used for model development and should not be serially correlated. For example, it would be preferred that only 1 hour of data from a given afternoon be used, since the data from the other hours on that afternoon are undoubtedly correlated with each other. Unfortunately, in the air quality business data sets are seldom large enough to make this possible.

Next the averaging period, sampling period, and time and space pairing of the data set must be chosen. The fractional bias and the NMSE can be calculated for all data paired in time and space (i.e., C_p-C_o is used from each monitor and each time). Weil and Brower (Reference 84) prefer to use the maximum observed and predicted concentrations at each averaging period on a given monitoring arc, but this method requires extensive monitoring over wide arcs. Or, FB and NMSE can be calculated for maximum concentrations at given monitoring locations independent of time. The combinations of data used are determined by the goals of the study.

Once the bias and NMSE are calculated for a number of models for the given data set, confidence limits should be calculated to answer the following questions:

1. Is the bias significantly different from zero?
2. Is the bias for Model i significantly different from the bias for Model j?
3. Is the NMSE for Model i significantly different from the NMSE for Model j?

Because the distributions of these parameters are not easily transformed to a normal distribution, standard analytical procedures for calculating confidence limits may not be accurate. The bootstrap resampling procedure described by Efron (Reference 85) is used instead. EPA scientists (Reference 82) has begun to apply the bootstrap procedure to air quality model studies, and many other examples are given by Hanna and Heinold (Reference 83).

The bootstrap procedure is best explained using a simple example. Suppose a set of 100 pairs of observed and predicted concentrations are available, and the averages $\bar{C}_o$ and $\bar{C}_p$ are calculated. In the bootstrap procedure, 100 new pairs are selected randomly (with replacement) by computer (or by hand, if you have the time) from the original set and new values of the averages $\bar{C}_o$ and $\bar{C}_p$ are calculated. This is done 100 to 1000 times (depending on the computer costs and availability), giving a histogram or pdf of $\bar{C}_p$-$\bar{C}_o$. If all of the resampled values of $\bar{C}_p$-$\bar{C}_o$ are, say, greater than zero, then it can be stated with great confidence that $\bar{C}_p$-$\bar{C}_o$ is significantly different from zero. But if the resampled $\bar{C}_p$-$\bar{C}_o$ distribution crosses zero at some point between the 5th and 95th percentiles, then it cannot be stated with 90 percent confidence that $\bar{C}_p$-$\bar{C}_o$ is significantly different from zero. This method of estimating the 5th and 95th percentile has been called the "seductive bootstrap" by some statisticians, who recommend the alternate procedure of using the calculated variance to determine these percentiles. However, we have found that the differences in these methods are minor for most air quality data.

An example of an application of this procedure is given in Figure 13, where seven models have been tested with the EPRI power plant data mentioned above and the distributions of the mean bias are drawn as Tukey "whisker-plots." It cannot be stated with 90 percent confidence that the mean bias, $\bar{C}_p$-$\bar{C}_o$ for models 2 and 5 is significantly different from zero. The $\bar{C}_p$-$\bar{C}_o$ for the other models is significantly different from zero by this criterion.

The bootstrap procedure can be applied to any statistic calculated from the original data set. While the use of the bias and the NMSE is emphasized above, the correlation, r, the median, or any other parameter could just as easily have been used. Even nonstandard statistics such as the

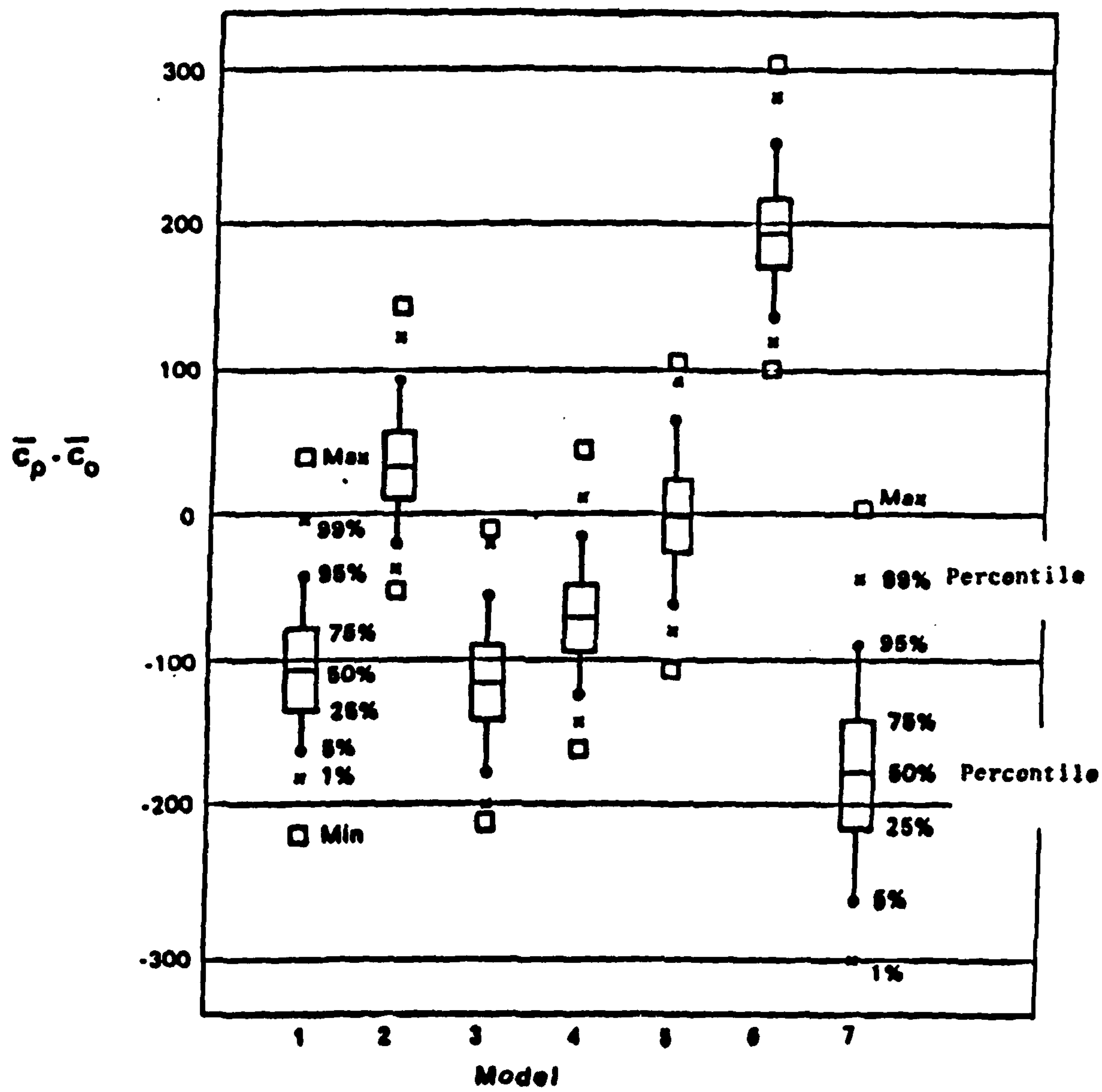


Figure 13. Illustration of "Whisker-Plots" of Cumulative Distribution Functions of $\bar{C}_p-\bar{C}_o$ for Seven Different Models, as Determined by Bootstrap Resampling. It is Concluded with Better than 95% Confidence that the Difference $\bar{C}_p-\bar{C}_o$ is not likely to be Zero for All Models Except Numbers 2 and 5, (i.e., the 5% or 95% Points on the Whisker Plots do not Overlap Zero).

percent of predictions within a factor of two of the observations can be used. A generalized algorithm using the bootstrap procedure to calculate confidence intervals on the statistics FB and NMSE has been written for the IBM PC microcomputer and is available on floppy disk from the authors.

C. INTERPRETATION OF CONFIDENCE INTERVALS

One of the primary advantages of our model evaluation procedure is its application of the concept of confidence intervals. Some analytical methods of calculating confidence limits are well-known. For example, if a "parent distribution" has a normal or Gaussian distribution with mean of 0.0 and standard deviation of sigma, and many samples of size n are randomly drawn from this distribution, then the calculated means of these samples will have an expected mean of 0.0 and standard deviation sigma/$\sqrt{n}$. This is called the central limit theorem.

If a calculated mean of a given sample of size n is M and the standard deviation is S, and n is greater than about 30, then it can be stated that there is a 95 percent chance that the mean of the parent distribution from which the sample was drawn is in the range from $M - 2S/\sqrt{n}$ to $M + 2S/\sqrt{n}$. This is called the Student t procedure, and is summarized in many texts including Panofsky and Brier (Reference 86).

In air pollution modeling, it is often asked if the model predictions are significantly different from the observations. For example, what are the chances that $\bar{C}_p - \bar{C}_o$ would equal zero for a given set of pairs of C_p and C_o? The analytical methods (e.g., central limit theorem) described above for estimating confidence limits are based on the presumption that the distribution of $C_p - C_o$ is normal or Gaussian. If it is not, the analytical methods begin to fail and should be replaced by a resampling procedure (such as the bootstrap) that does not care whether the distribution is Gaussian or not. Section IV.B recommends the "seductive bootstrap" for resampling air quality data, but the "jackknife" or the "multihalver" could be used (Reference 87). The output of any of these methods is an estimate of the mean, M, and the standard deviation, S, from which the 95 percent confidence interval equals (as before) $M - 2S/\sqrt{n}$ to $M + 2S/\sqrt{n}$ (for $n > 30$).

For example, suppose $M = \bar{C}_p - \bar{C}_o = 10\ \mu g/m^3$, $S = 40\ \mu g/m^3$, and $n = 100$.

Then the 95 percent confidence interval on M ranges from 2 $\mu g/m^3$ to 18 $\mu g/m^3$. In this case the confidence interval does not include zero and the interpretation is that the mean bias $\bar{C}_p - \bar{C}_o$ is significantly different from zero (with 95 percent confidence). This conclusion would be the opposite if $M = \bar{C}_p - \bar{C}_o$ were reduced to 5 $\mu g/m^3$.

The same approach is applied to differences between models, where $M = \bar{C}_{p1} - \bar{C}_{p2}$, with subscripts 1 and 2 representing two different models. Experience shows that significant differences between models occur less frequently than intuition would suggest, as will be seen in the next section.

SECTION V

PRELIMINARY APPLICATIONS OF MODEL EVALUATION PROCEDURES

The model evaluation procedures discussed in Section IV were applied to predictions of several models for several types of field experiments. Four specific applications are described in this section.

A. DENSE GAS MODEL COMPARISONS

Alp et al. (Reference 88) presented Table 13, which contains predictions of four dense gas models for six runs during the Maplin Sands LPG experiments. We did not include the FEM3 model in our comparisons, since it did not make predictions for two of the runs. Table 14 contains the results of the model evaluation, showing that the fractional bias FB and the normalized mean square error NMSE are quite small for all models. In fact, FB is not significantly

TABLE 14. CONCLUSIONS FROM PROPANE RUNS REPORTED BY ALP (REFERENCE 75)

MODELS:	COBRA III	HEGADAS II	DEGADIS
FB	.002	0.09	.151
NMSE	.09	.09	.10

MODELS FOR WHICH FB IS NOT SIGNIFICANTLY DIFFERENT FROM 0 (AT 95 PERCENT CONFIDENCE LEVEL): ALL

MODELS FOR WHICH DFB IS NOT SIGNIFICANTLY DIFFERENT FROM 0 (AT 95 PERCENT CONFIDENCE LEVEL): COBRA III - HEGADAS II.

MODELS FOR WHICH DNMSE IS NOT SIGNIFICANTLY DIFFERENT FROM 0 (AT 95 PERCENT CONFIDENCE LEVEL): ALL

Note $FB = (\bar{C}_o - \bar{C}_p)/0.5\ (\bar{C}_o + \bar{C}_p)$

$$NMSE = \overline{(C_p - C_p)^2}/\bar{C}_o\ \bar{C}_p$$

TABLE 13. SUMMARY OF RESULTS FOR PROPANE RUNS (REFERENCE 75).

		HAZARD ZONE LENGTH . . . (M)				
Run No.	U_o (m/s)	Maplin Sands (obs. to LFL)*	COBRA III Prediction (2.1%)	HEGADAS II Prediction (2.1%)	DEGADIS Prediction (2.1%)	FEM3 Prediction (2.1%)
43	5.5	215 ± 20	257	244	218	-
46	8.1	230 ± 35	227	225	195	-
47	5.6	235 ± 25	341	322	290	220
49	6.2	270 ± 25	205	211	190	140
50	7.9	220 ± 40	279	278	240	190
54	3.8	450 ± 70	308	294	260	270

LFL: Lower Flammability Limit = 2.1%

different from zero (95 percent c.i.) for any of the models. However, FB for the DEGADIS model is significantly different from FB for the other two models. The three NMSE values are not significantly different from each other. It is concluded that all three models perform well and that their performance is not significantly different. The broad confidence limits are a result of the relatively small number of data that were used (n = 6).

As another example, Figure 14 shows predictions of four models for the Thorney Island Trial 14 freon experiment (from Reference 49). One model (Eidsvik) appears to be greatly in error, and the other models have much less error. Concentrations at five times after release were considered. The output of the statistical software is given in Table 15. The Eidsvik model yields a fractional bias FB that is significantly different from zero (95 percent c.i.) and is different from the FB of the other models. The NMSE for the four models are not significantly different. This apparent lack of difference in NMSE among models is partly due to the small sample size (n = 5 in this case). The smaller n, the larger the confidence interval, as described in Section IV C. Generally about 100 separate field trials are needed to show significant differences among similar models. It can be concluded from this application that the performance of all four models is similar, with the possible exception of the Eidsvik model as an outlier.

TABLE 15. CONCLUSIONS FROM THORNEY ISLAND TRIAL 14 DATA.

MODELS:	COX,	EIDSVIK,	HEGADAS,	MARIAH
FB	-.27	-1.36	.17	-.29
NMSE	.13	.28	.34	.13

MODELS FOR WHICH FB IS NOT SIGNIFICANTLY DIFFERENT FROM 0 (AT 95 PERCENT CONFIDENCE LEVEL): ALL BUT EIDSVIK.

MODELS FOR WHICH DFB IS NOT SIGNIFICANTLY DIFFERENT FROM 0 (AT 95 PERCENT CONFIDENCE LEVEL): COX-HEGADAS, COX-MARIAH, HEGADAS-MARIAH.

MODELS FOR WHICH DNMSE IS NOT SIGNIFICANTLY DIFFERENT FROM 0 (AT 95 PERCENT CONFIDENCE LEVEL): ALL.

Note $FB = (\bar{C}_o - \bar{C}_p)/0.5\ (\bar{C}_o + \bar{C}_p)$ $NMSE = \overline{(C_p - C_p)^2}/\bar{C}_o\ \bar{C}_p$

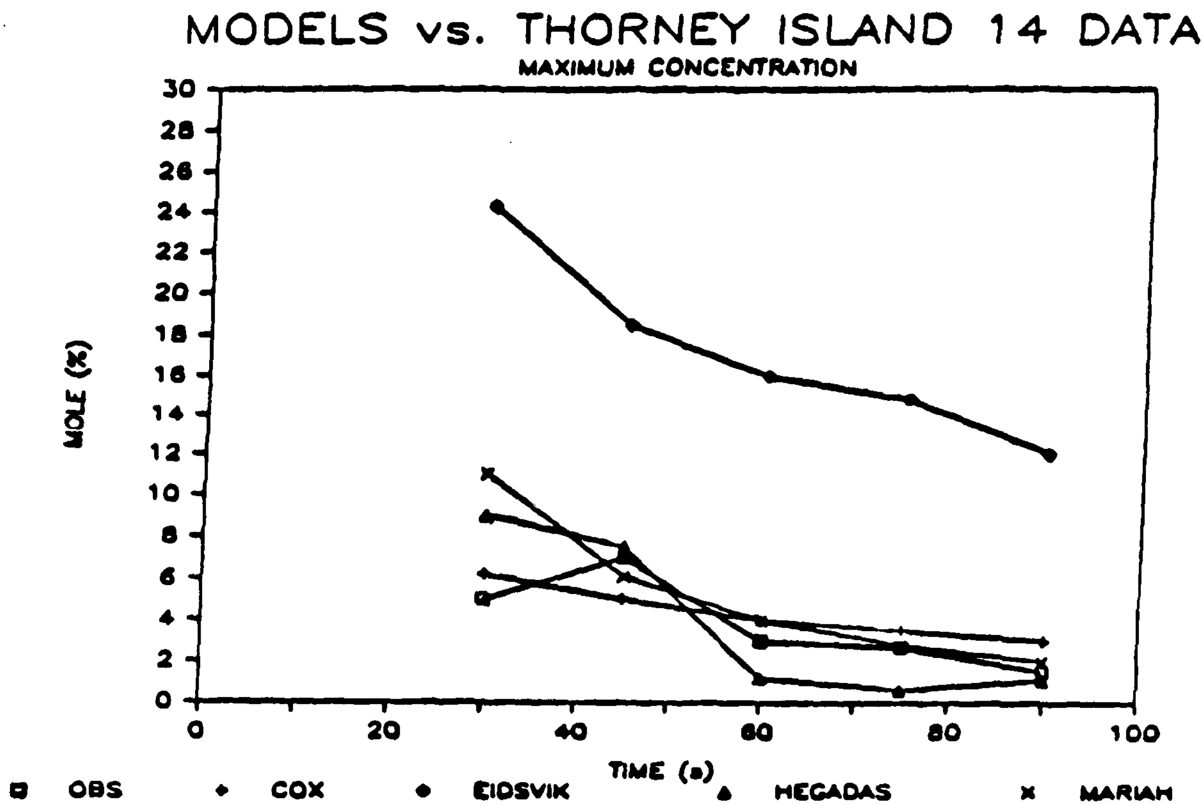


Figure 14. Models vs. Thorney Island 14 Data.

B. AFTOX OB/DG COMPARISONS

The new AFTOX model (Reference 2) is intended to replace the 20-year old OB/DG model, which has been used by the U.S. Air Force to calculate concentrations and hazard corridors for many years. It is described in more detail in Section II. But is the AFTOX model a significant improvement over the OB/DG model? Kunkel (Reference 2) tested both models using the Prairie Grass, Ocean Breeze, Dry Gulch and Green Glow data, and kindly supplied us with meteorological data and model predictions for each experiment. A summary of the tests is given in Table 16.

TABLE 16. SUMMARY OF INERT TRACER TESTS.

Test	Trials	Arc Distances	Conditions
Ocean Breeze Cape Canaveral	76 FP	1.2, 2.4, 4.8 km	Daytime
Dry Gulch Vandenberg	109 FP	Course 1: 2, 3, 5.7 km Course 2: 0.85, 1.5, 4.7 km	Daytime
Green Glow Hanford	66 FP	0.2, 0.8, 1.6, 3.2, 12.8, 25.6 km	Stable
Prairie Grass Nebraska	70 SO_2	0.05, 0.1, 0.2, 0.4, 0.8 km	Daytime

This evaluation is hampered by the fact that the OB/DG model is in fact an empirical model derived from the Ocean Breeze, Dry Gulch, and Prairie Grass data. The model developers (Reference 27) did separate the data into a developmental data set used for deriving the OB/DG equation and a test data set used for its evaluation, but the test data are not independent in a true statistical sense. The Green Glow data are independent, but are weighted towards stable conditions, for which the OB/DG model is not calibrated.

Our model evaluation results are summarized in Table 17, where the number in the experiment code indicates a given monitoring arc, and n equals the

TABLE 17. STATISTICAL EVALUATION FOR PRAIRIE GRASS (PG), GREEN GLOW (GG), OCEAN BREEZE (OB), AND DRY GULCH (DG) EXPERIMENTS. THE NUMBER FOLLOWING THE SITE CODE IS THE MONITORING ARC.

EXPER	n	FB AFTOX	FB OBDG	NMSE AFTOX	NMSE OBDG	DOES D FB=0	DOES DNMSE=0
pg1	52	-0.27	-0.34	0.49	0.49	yes A**	yes
pg2	52	-0.03*	-0.11	0.005	0.098	yes A	yes A
pg3	52	0.087	-0.013*	0.062	0.066	no 0	yes A
pg4	52	0.031*	-0.01*	0.14	0.092	yes 0	yes 0
pg5	52	-0.16*	-0.11*	0.48	0.42	yes 0	yes 0
gg1	24	0.06*	-0.64	0.13	0.83	no A	no A
gg2	24	0.15*	-0.45	0.17	0.54	no A	no A
gg3	24	-0.13*	-0.5	0.39	0.56	no A	yes A
gg4	24	-0.4	-0.5	0.68	0.57	yes A	yes A
gg5	24	-0.8	-0.33	4.1	0.68	no 0	yes 0
gg6	24	-0.91	-0.1*	5.5	0.5	no 0	no 0
ggall	144	-0.44	-0.44	2.74	0.63	no	no 0
ob1	67	-0.1*	-0.01*	0.39	0.3	yes 0	yes 0
ob2	67	-0.37	-0.09*	0.9	0.63	no 0	no 0
oball	134	-0.24	-0.05	0.68	0.47	yes 0	no 0
dgb1	45	-0.07*	-0.11*	0.31	0.2	yes A	yes 0
dgb2	45	0.2*	0.3	0.43	0.23	yes A	yes 0
dgball	90	0.058	0.077	0.36	0.21	yes A	yes 0
dgd1	51	-0.08*	-0.33	0.22	0.3	no A	yes A
dgd2	51	-0.26	-0.36	0.36	0.3	yes A	yes 0
dgdall	102	-0.18	-0.35	0.3	0.3	yes A	yes

* Indicates that FB is not significantly different from 0.0 (95 percent c.i.)

** Indicates which model had a lower FB or M (A=AFTOX and 0=OBDG)

number of data. There is not a clear distinction between the models, with the AFTOX model performing better on about as many data sets as the OB/DG model. It is difficult to decide how to combine all the various results in the table. In fact, if one counts only those data sets where one model showed a significantly better FB or NMSE, then the "score" is AFTOX 11 to OB/DG 13. It is concluded that the AFTOX and OB/DG models are not significantly different when compared to these data sets.

The OB/DG model is known (Reference 30) to underpredict (negative FB) for stable conditions, and the AFTOX model was developed with the intent of correcting this deficiency. However, looking at the results from the six Green Glow monitoring arcs, it is seen that an improvement has been made only at the inner three monitoring arcs. By the fourth arc, the models each underpredict by about 50 percent, and by the sixth arc, the AFTOX model underpredicts by 90 percent and the OB/DG model is fairly close (underprediction of only 10 percent). Kunkel (Reference 2) also mentions these problems. Further enhancements of the AFTOX model are needed for these large downwind distances.

C. CHARM APPLIED TO FLAT TERRAIN EXPERIMENT

As part of the model evaluation procedure, the Air Force toxic model, AFTOX, and the CHARM heavy gas diffusion model, were tested using the Ocean Breeze passive tracer tests. CHARM was run "as is" (a necessity since the code is proprietary). All data were input manually. Since the Ocean Breeze study used zinc sulfide as the tracer and CHARM did not model this substance, ethane was chosen from the chemical list because it is fairly nonreactive and close to the molecular weight of air. Model inputs were as close to the actual conditions as possible. It was thought that a good test of CHARM's ability to model passive releases would result from using ethane with no appreciable release speed. The model was allowed to choose the stability class. CHARM chose Pasquill-Gifford class C; class D was reported in Ocean Breeze.

The "user supplies description" option was used to characterize the release. The tracer storage pressure was set slightly higher than the ambient pressure to supply the needed force for the release to occur. The release was continuous in nature with a constant emission rate. The release was of 30

minutes duration and concentration values were output for the downwind distances of interest. Comparisons of CHARM and AFTOX predictions with the measurements for Runs 2 and 4 of Ocean Breeze show that AFTOX is much closer to the measurements than CHARM but that both models overpredict. CHARM is seen to overpredict by a factor of two to four (see Table 18). Confidence limits were not calculated since the number of data is so small.

TABLE 18. COMPARISON OF AFTOX AND CHARM USING DATA FROM PROJECT OCEAN BREEZE

RUN #	DISTANCE (m)	MEASURED (mg/m^3)	AFTOX (mg/m^3)	CHARM (mg/m^3)
2	1207	0.0095	0.01	0.043
2	2414	0.023	0.037	0.064
4	1207	0.011	0.013	0.043
4	2414	0.0031	0.0048	0.0064

D. *AFTOX, SAFER, AND TRACE MODELS APPLIED TO THORNEY ISLAND TRIAL 7 DATA*

Data from the Thorney Island dense gas experiments were also used in the model evaluation tests. In our application of the AFTOX model, Thorney Island Trial 7 was modeled as an instantaneous release of Freon-12, which was included in the AFTOX chemical list. Output option 3 was chosen and used to determine the maximum concentration and the distance downwind at which it occurred at various times. Comparisons of AFTOX predictions with those of SAFER and TRACE are shown in Table 19. All of the models overpredict; SAFER by a factor of 2 for both distances shown, TRACE by a factor of 3, and AFTOX by a factor of 5. The reason for the overprediction of the AFTOX model (which considers the tracer cloud to be non buoyant) with respect to the SAFER and TRACE models (which account for dense gas effects) is that the AFTOX model assumed that the puff is always transportd by the wind speed. In contrast, the SAFER and TRACE models assume that the puff is only slowly accelerated up to the full wind speed. Even though the puff volume increases more rapidly with time in the AFTOX model, the AFTOX puff reaches a given downwind distance much faster than the SAFER or TRACE puffs.

TABLE 19. MODEL COMPARISON USING DATA FROM THORNEY ISLAND TRIAL 7.

DISTANCE (m)	MEASURED (ppm)	AFTOX (ppm)	SAFER (ppm)	TRACE (ppm)
240	7000	33000	12000	19000
407	3800	11508	6900	7600

SECTION VI

CONCLUSIONS AND RECOMMENDATIONS

This preliminary study has touched upon a large number of issues related to the estimation of hazard response model uncertainty. The results of this research are summarized below and recommendations for further research are given.

A. CONCLUSIONS FROM PRELIMINARY APPLICATION

The intent of this project was to develop and test a quantitative method for assessing the uncertainty of hazard response models. At the beginning of the project it was not obvious that it would be possible to develop a generalized framework for this purpose. However, the different components have pulled together reasonably well and there is hope for a satisfactory completion of Phase II. The specific conclusions listed below follow the outline of this report.

1. Literature Review; Acquisition of Models, Data Sets, and Model Predictions.

The available literature on hazard response models, evaluations, and field studies was reviewed in order to develop an awareness of previous work on this subject. Previous work on hazard response model evaluations had been limited to simple comparisons and did not account for confidence limits on the performance measures calculated. However, a more complete framework was available from other reports on air quality model evaluations for the EPA and others. Consequently, procedures do exist for accounting for confidence limits, stochastic fluctuations, data input uncertainties, and model physics errors.

Air Force hazard-response model evaluation reports were reviewed, including the older OB/DG model reports, the newer AFTOX model development, the N_2O_4 experiments, the dense gas model development by Raj, the data review by Ermak, the model sensitivity studies by Carney, the revisions of DEGADIS by Spicer and Havens, the applications of CHARM, and the model

comparisons by Key and his coworkers.

Hazard response models were collected, installed on our microcomputers, and tested. These models include OB/DG, AFTOX, CHARM, OME, INPUFF, AVACTI-II, MADICT, RVD, D2PC, SPILLS and SLAB.

Sets of predictions by various models for various data sets were collected and archived. These include the OB/DG and AFTOX predictions for the Prairie Grass, Green Glow, Ocean Breeze and Dry Gulch experiments, and several dense gas model (for example, SAFER, CHARM, DEGADIS) predictions for data sets such as Thorney Island and Maplin Sands.

Comprehensive data sets for several experiments were obtained and archived on our microcomputers. These data sets include source terms and observed meteorological parameters for each time period during the experiment.

2. Analysis of Components of Model Uncertainty

A method for decomposing the total model uncertainty or mean square error into its three components (stochastic fluctuations, data input errors, and model physics errors) was outlined.

Stochastic concentration fluctuations are seen to be the result of stochastic turbulent fluctuations in the atmosphere. The variance of these fluctuations can be predicted using analytical methods available from the literature. The ratio of the standard deviation due to the stochastic fluctuations to the mean of the observed concentrations ranges from about 0.1 to 1.0 for observations with short-term averaging (on the order of one second). A formula for expressing this ratio as a function of averaging time is proposed and some examples given of its application.

Data input errors are significant for standard Air Force meteorological instrumentation. Typical instrument errors are about 10 percent at a minimum. Even in research-grade experiments these instrument errors are typically 5 to 10 percent. These errors can be accounted for in analytical procedures for assessing model sensitivity to input variability.

Model physics errors are difficult to estimate, since they equal the difference between the total model error and the sum of the stochastic and the data input error components. A method was developed for calculating the contribution of various model components to the uncertainty. This method involves the calculation of the total variance of the model residuals for various combinations of model components (for example, the following combinations were tested: Q, 1/u, Q/u, Q/ux, $Q/(\sigma_\phi \sigma_\theta x^2)$, etc.). It is found that for the Prairie Grass field data sets, the combination $(\sigma_\phi \sigma_\theta x^2 u)$ accounts for most of the variance, and additions of further 'improvements' in model physics do not reduce the total variance any more.

3. Framework of Model Evaluation Procedure

A quantitative model evaluation procedure was derived and a software package was written and tested. This procedure assumes that a table exists that contains a listing of observations of concentrations or hazard corridor lengths, along with predictions from one or more models of the same quantity. The fundamental model performance measures that are suggested are the fractional bias $FB = (\bar{C} - \bar{C}_p)/0.5*(\bar{C}_o + \bar{C}_p)$ and the normalized mean square error $NMSE = \overline{(C_o - C_p)^2}/\bar{C}_o\bar{C}_p$.

Confidence limits are calculated for FB to determine whether it is significantly different from zero (that is, are the predictions of the model significantly different from the observations?). The bootstrap resampling procedure is used to calculate the confidence limits. In addition confidence limits on differences in FB and NMSE between models are calculated to determine whether the predictions of the models are significantly different.

4. Results of Preliminary Application of Model Evaluation Procedure

The model evaluation procedure described above was applied to four sets of model predictions and observations:

a. Dense gas model comparisons.

A limited set of data was available from the Maplin Sands and the Thorney Island dense gas dispersion experiments. Model predictions were also available. Because the data sets are so small (n = 4 or 5), the confidence limits on FB and NMSE are so large that there is generally no significant difference among the models, although visual inspection of the predictions would suggest that one model appears to perform better than the others.

b. AFTOX and OB/DG model comparisons.

These models are intended for application to non buoyant sources, and predictions were available for the Prairie Grass, Green Glow, Ocean Breeze, and Dry Gulch data sets. In each case the sample size is about 50 to 100. The comparison is hampered by the fact that the OB/DG model was actually derived from some of these data sets. In most cases, the two model predictions are significantly different from each other, although there is not a clear trend towards one model or another. The OB/DG model underpredicted the stable Green Glow data, but the AFTOX model shows an improvement over the OB/DG model for the Green Glow data set only at monitoring arcs close to the source. At the farthest monitoring arcs the OB/DG model predictions are significantly better than the AFTOX model predictions.

c. CHARM model applied to non-buoyant source.

The CHARM model, which accounts for a wide variety of source types (including dense gases) was applied to the Prairie Grass experiment, where the source was non buoyant. It is found that the CHARM model is conservative by a factor of about two to four. These predictions are significantly different from those of the AFTOX or OBDG models.

d. AFTOX model applied to dense gas experiment.

The AFTOX model does not apply to dense gas sources. However, it was applied to a Thorney Island dense gas run to determine the typical error that should be expected. It is found that the model

overpredicts by a factor of four or five. Apparently the overprediction by the AFTOX model is due to the fact that it allows the puff to immediately assume the wind speed and thus be transported more quickly to the monitor locations.

B. RECOMMENDATIONS FOR FURTHER STUDY.

The preliminary conclusions listed above suggest that a more extensive research program would be worthwhile. It is recommended that this research program contain the following components:

1. Archiving of Information

Much information on data sets and model predictions was acquired under Phase I of this project. A comprehensive search for further data sets and model predictions should be undertaken and the data sets installed in our microcomputers. This would include all hazardous gas data summarized by Ermak (Reference 40) and recent field tests using HF and other toxic gases. In addition, sets of model predictions would be acquired from model developers for as many of these experiments as possible. Similar data would be obtained for non buoyant experiments (for example, Cabauw) and model predictions.

To complete the tables of model predictions and observations, models will be run for the missing periods. Several of these models are on-hand. It is hoped that the following models can be acquired: Raj's dense gas model developed for the Air Force (ADAM), the PC version of DEGADIS developed for the API, a PC version of SLAB developed for the Air Force and for the API, and a PC version of CAMEO. Also it is possible that some models that are currently proprietary will be released to the public (for example, SAFER, EAHAP).

2. Determination of Uncertainty Components.

A start was made in determining the components of model uncertainty. In the future, improved estimates of typical concentration fluctuations for hazardous gases can be made using data from field experiments and using theoretical concepts suggested by, for example, Chatwin (Reference 74). A better estimate of the integral time and distance scales must be made

in order to calculate the effects of averaging times and distances. We plan to use analytical formulas developed by Hanna (Reference 89) under an Army Research Office contract.

The Air Force does not have a good idea of the data uncertainties in their meteorological instruments. It is desirable to obtain better estimates of this uncertainty through a field program in which (1) two similar instruments are set up at the same location; (2) a high-quality baseline instrument is set up next to the Air Force instrument; (3) several similar instruments are set up along a path at separations of 10 meters, 20 meters, 50 meters, 100 meters, and so on; (4) wind tunnel and environmental chamber tests of instruments are made; and (5) manufacturer QA/QC procedures are carefully reviewed. The impact of individual input data errors on total model uncertainty can be investigated using analytical methods suggested and tested by Carney (Reference 5).

Further application of the variance reduction analysis should take place using the comprehensive data sets and model prediction tables. In this way it can be determined whether a given model (for example, AFTOX or SAFER) yields any improvement over simple models such as $C = Q/u$. The expected improvement in the model due to the inclusion of new model physics parameters can therefore be estimated.

3. Development of Model Evaluation Procedures.

The model evaluation procedures described above are reasonable and can produce estimates of confidence limits. Future studies should test further candidate model performance measures, such as the correlation coefficient and the fraction of predictions within a factor of two of the observations. The recent review by Ermak and Merry (Reference 6) will also be useful in devising more reasonable performance measures. In addition, alternate resampling schemes should be tested, such as the jackknife and the multihalver (Reference 87). The accuracy of these schemes can be tested using concocted data sets from a known parent distribution (for example, the normal or log-normal distribution).

4. Application of Procedure.

A few applications of the preliminary model evaluation software have been made and suggest that it is possible to arrive at conclusions regarding the accuracy of a given model or whether two model predictions are significantly different from each other. In the future the revised model evaluation procedures should be tested with the comprehensive data sets to be acquired. Models of interest to the Air Force (ADAM, AFTOX, CHARM, DEGADIS, SLAB, and OB/DG) will be included in the analysis. After this application is complete, we will have a good idea of the typical accuracy of hazard response models and whether any one model is better than another. We note that it is important that these decisions should be made using independent data sets (that is, data sets that were not used in the development of the model).

5. Software Package.

The final product should be a user-friendly software package that can easily be used to assess the uncertainty of hazard response models. This package will be easily transportable from one microcomputer to another and can be understood and operated by Air Force scientists and engineers.

6. Confidence Levels of Models.

Once experience is gained with the model evaluation methodology, it will be possible to provide decisionmakers with models that include confidence intervals as well as basic calculations.

REFERENCES

1. Hanna, S.R. and Drivas, P.J. Guidelines for Vapor Cloud Dispersion Modeling. AIChE/CCMS, New York, May 1987.

2. Kunkel, B.A. AFTOX Evaluation, AFGL Draft Report, 1987.

3. Kunkel, B.A. A Comparison of Evaporative Source Strength Models for Toxic Chemical Spills, AFGL-TR-83-0307, AFGL, Hanscom AFB, MA 01731, 1983.

4. Kunkel, B.A. Development of an Atmospheric Diffusion Model for Toxic Chemical Releases, AFGL-TR-85-0338, AFGL, Hanscom AFB, MA 01731, 1985.

5. Carney, T.A. A Comparative Study and Evaluation of Four Atmospheric Dispersion Models with Present or Potential Utility in Air Force Operations, and a Feasibility Study and Test Application of Uncertainty Estimates to an Atmospheric Dispersion Model with Final reports for AFESC Mini-Grant Program, Tyndall AFB, FL 32403-6001, 1986 and 1987.

6. Ermak, D.L. and Merry, M.H. A Methodology for Evaluating Heavy Gas Dispersion Models. Draft Final Report for AFESC, Tyndall AFB, Florida, February 1988.

7. Spicer, T.O., Havens, J.A., Tebeau, P.A., and Key, L.E. DEGADIS - A Heavier than Air Gas Atmospheric Dispersion Model Developed for the U.S. Coast Guard. APCA Annual Meeting Paper 86-42.2, 1986.

8. Key, L.E. and Bowman, C.R. "Hypergolic Accidental Release Model." Preprints, Fifth Joint Conf. on Appl. of Air Poll. Meteorol., AMS, pp 307-310, 1986.

9. Whitacre, C.G., Griner, J.H., Myirski, M. and Sloop, D.W. Personal Computer Program for Chemical Hazard Prediction (D2PC). U.S. Army CRDC Tech Report CRDEC-TR-87021, CRDC, Aberdeen Proving Ground, MD 21010-5423, 1986.

10. Layland, D.E., McNaughton, D.J. and Bodner, P.M., "Assessing the Reliability of Dispersion Models for Hazardous Materials Releases." Proc. 5th Joint Conf. on Applic. of Air Poll. Meteorol. AMS, 1986.

11. Paine, R.J., Pleim, J.E., Heinold, D.W., and Egan B.A., "Physical processes in the release and dispersion of toxic air contaminants." Paper 86-76.3 at 79th Ann. Meeting of the Air Poll. Control Assoc., Minneapolis, 1986.

12. Heinold, D.W., Paine, R.J., Walker, K.C., and Smith D.G., "Evaluation of the AIRTOX dispersion algorithms using data from heavy gas field experiments." Proc. 5th Conf. on Application of Air Poll. Meteorol., AMS, Boston, 1986.

13. Ermak, D.L., Chan, S.T., Morgan, D.L., and Morris, L.K. "A Comparison of Dense Gas Dispersion Model Simulations with BURRO series LNG spill test results." Journal of Hazardous Materials, 6, pp. 129-160, 1982.

14. Ermak, D.L. and Chan, S.T., "A Study of heavy gas effects on the atmospheric dispersion of dense gases." Paper VI.6, Proc. of 15th Int. Tech. Meeting on Air Poll. Modeling and its Applic., NATO/CCMS, 1985.

15. McRae, T.G. Analysis and Model/Data Comparisons of Large-Scale Releases of Nitrogen Tetroxide. ESL-TR-85-06, AFESC, Tyndall AFB, FL, June 1985.

16. Alp, E., Alp, S., Brown, P.M., and Portelli, R.V., "Heavy gas dispersion modeling." Paper VI.2, Proc. of 15th Int. Tech. Meeting on Air Poll. Modeling nad its Applic., NATO/CCMS, 1985.

17. Riou, Y.J. and Saab, A.E., "A three-dimensional numerical model for the dispersion of heavy gases over complex terrain." Paper VI.7, Proc. of 15th Int. Tech. Meeting on Air Poll. Modeling and its Applic., NATO/CCMS, 1985.

18. Balentine, W., and Eltgroth, W. Validation of a hazardous spill model using N_2O_4 and LNG spill data. Presented at the 78th Annual Meeting of the Air Pollution Control Association, Detroit, Michigan, 1985.

19. Lewellen, W.S., Sykes, R.I., and Parker, S.F., Comparison of the 1981 INEL Dispersion Data with Results from a Number of Different Models. NUREG/CR-4159 NRC, Washington, DC, 1985.

20. Wheatley, C.J., Brighton, P.W.M., and Prince, A.J., "Comparison between data from the heavy gas dispersion experiments at Thorney Island and predictions of simple models," Paper VI.5, Proc. of 15th Int. Tech. Meeting on Air Poll. Modeling and its Applic., NATO/CCMS, 1985.

21. Wheatley, C.J., "Factors Affecting Cloud Formation from Releases of Liquified Gases." I. Chem. E. Symp. on Refinement of Estimates of the Consequences of Heavy Toxic Vapour Releases, 1986.

22. Puttock, J.S., Colenbrander, G.W., and Blackmore, D.R., "Maplin Sands 1980: Dispersion Results from Continuous Releases of Refrigerated Liquid Propane and LNG." Air Poll. Modeling and its Applic. (C. DeWispelaere, ed.), Plenum, pp 353-373, 1984.

23. Puttock, J.S. and Colenbrander, G.W., "Thorney Island Data and Dispersion Modeling." J. Haz. Mat., 11, pp 381-397, 1985.

24. Fay, J.A. and Ranck, D.A., "Comparison of Expeiments on Dense Gas Cloud Dispersion." Atmos. Environ., 17, pp 239-248, 1983.

25. Havens, J.A. and Spicer, T.O., Development of an Atmospheric Dispersion Model for Heavier-than-Air Gas Mixtures. Report No. CG-D-22-85 for the U.S. Coast Guard by the Chem. Eng. Dept., University of Arkansas, Fayetteville, Ark. 72701, Three Volumes, 1985.

26. Spicer, T.O. and Havens, J.A. Development of Vapor Dispersion Models for Nonneutrally Buoyant Gas Mixtures - Analysis of USAF/N_2O_4 Test Data. ESL-TR-86-24, AFESC, Tyndall AFB, FL 32403, 1986.

27. Nou, J.V. The Ocean Breeze and Dry Gulch Diffusion Programs, AFCRL, Hanscom AFB, MA, 1963.

28. Haugen, D.A. (Ed.), "Project Prairie Grass, A Field Program in Diffusion", Geophysical Research Papers, No. 59, Vol. III, Report AFCRC-TR-58-235 (ASTIA Document AD-217076). Air Force Cambridge Research Center, 1959.

29. Haugen, D.A., and Fuquay, J.J. The Ocean Breeze and Dry Gulch Diffusion Programs. Vol. I, USAEC Report HW-78435, Air Force Cambridge Research Laboratories, Mass., and Hanford Atomic Products Operation, Washington, 1963.

30. Ohmstede, W.D., Dumbauld, R.K. and Worley, G.G. Ocean Breeze/Dry Gulch Equation Review. ESL-TR-83-05, Engineering and Services Laboratory, Air Force Engineering and Services Center, Tyndall AFB, Fla., December 1983.

31. Kunkel, B.A., An Evaluation of the Ocean Breeze/Dry Gulch Dispersion Model. AFGL-TR-84-0313, AFGL, Hanscom AFB, MA 01731, 18 pp. 1984.

32. Clewell, H.J., Estimation of Hazard Corridors for Toxic Liquid Spills, 1980 JANAF Propulsion Meeting, AFESC, 1980.

33. Kahler, J.P., Curry, R.G., and Kandler, R.A., Calculating Toxic Corridors, AWS/TR-80/003, Air Weather Service (MAC), Sccott AFB, IL, 92 pp., 1981.

34. Smith, F.B. "A Scheme for Estimating the Vertical Dispersion of a Plume from a Source Near Ground Level." Proc. of the Third Meeting of the Expert Panel on Air Pollution Modeling, NATO Committee on the Challenges of Modern Society, Paris, France, 1972.

35. RADIAN Corp., Emergency Modeling of Hazardous Air Pollutants. ESL-TR-84-55, AFESC, Tyndall AFB, FL, 1985.

36. Spicer, T.O., Havens, J.A., and Key, L.E. "Extension of DEGADIS for Modeling Aerosol Releases." Int. Conf. on Vapor Cloud Modeling. AICHE, New York, pp 416-438, 1987.

37. Freeman, D.L., Egami, R.T., Robinson, N.F., and Watson, J.G. "A Method for Propogating Measurement Uncertainties Through Dispersion Models." J. Air Poll. Control Assoc., 36, pp. 246-253, 1986.

38. Hanna, S.R. "A review of air quality model evaluation procedures." WMO Int. Conf. on Air Poll. Modeling and Its Applic., Leningrad, USSR, May 1986.

39. Raj, P.K., Venkatramana, M., and Morris, J. "A Hybrid Box-Gaussian Dispersion Model Including Considerations of Chemical Reactions and Liquid Aerosol Effects." Int. Conf. on Vapor Cloud Modeling. AICHE, New York, pp 453-494, 1987.

40. Ermak, D. L., Chapman, R., Goldwine, C., Goveia, J., and Rodean, H.C. Heavy Gas Dispersion Test Summary Report. Draft Report for AFESC, 1987.

41. Mercer, A. CEA/UKAEA Exchange Agreement on Heavy Vapor Dispersion Model Comparisons, 1987.

42. Venkatram, A., "Model Evaluation," in a Short Course on Air Pollution Modeling, AMS, Boston, 1986.

43. Mercer, A. "Methods of Validation of Heavy Gas Dispersion Models: A Review." IMA Conf. on Stably Strat. Flows and Dense Gas Dispersion, April 1986.

44. McNaughton, J., Atwater, A., Bodner, M., and Worley, G. Evaluation and Assessment of Models for Emergency Response Planning. TRC report prepared for the Chemical Manufacturers Association, 229 pp. 1986.

45. McNaughton, D.J., Worley, G.G. and Bodner, P.M., "Evaluating Emergency Response Models for the Chemical Industry." Chem. Eng. Progress, pp. 46-51, January 1987.

46. Lamb, R.G., "Air Pollution Models as Descriptions of Cause-Effect Relationships." Atmos. Environ., 18, pp 591-606, 1984.

47. Puttock, J.S., "Comparisons of Thorney Island Data with Predictions of HEGABOX/HEGADAS." J. Haz. Mat., 1987.

48. Wheatley, C.J., Prince, A.J. and Brighton, D.W.M., Comparison between Data from the Thorney Island Heavy Gas Trials and Predictions of Simple Dispersion Models, SRD R 355, UKAEA, Culcheth, UK, February 1986.

49. Energy Analysts, Inc. Validation of Heavy Gas Dispersion Models with Experimental Results of Thorney Island Trials. Report 85-72-140 for the API, Washington, DC, 1985.

50. Chikhliwala, E.D., Oliver, M., and Shinde, A.N. "Comparison of an Accidental Release Modeling Technique with Recent Spill Test Data." APCA Annual Meeting Paper 87-102.6, 1987.

51. Blewitt, D.N., Yohn, J.F. and Ermak, D.L. "An Evaluation of SLAB and DEGADIS Heavy Gas Dispersion Models Using the HF Spill Test Data." Int. Conf. on Vapor Cloud Modeling. AICHE, New York, pp 56-80, 1987.

52. Koopman, R.P., McRae, T.G., Goldwire, H.C., Ermak, D.L., and Kansa, E.J., "Results of Recent Large-Scale NH_3 and N_2O_4 Dispersion Experiments." 3rd Symp. on Heavy Gases and Risk Assessment, November 1984.

53. Zannetti, P. "A New Mixed Segment-Puff Approach for Dispersion Modeling." Atmos. Environ.,20, pp. 1121-1130, 1986.

54. Ludwig, F.L. Users' Guide for a Model of Airflow and Diffusion in Complex Terrain (MADICT), prepared for U.S. Army Research Office by SRI Int., Menlo Park, CA, 1985.

55. Ontario Ministry of the Environment. Portable Computing System for Use in Toxic Gas Emergencies, Report No. ARB-162-83-ARSP, 1983.

56. Petersen, W.B. and Lavdas, L.G. INPUFF 2.0 - A Multiple Source Gaussian Puff Dispersion Algorithm. Users Guide, ASRL, USEPA, Research Triangle Park, NC, 1986.

57. Fleischer, M.T. SPILLS, An Evaporation/Air Dispersion Model for Chemical Spills on Land. Shell Development Center, Westhollow Research Center, P.O. Box 1380, Houston, TX, 77001, 1980.

58. Draxler, R.R. "Diffusion and Transport Experiments." Chapter 9 in Atmospheric Science and Power Production. DOE/TIC-27601. USDOE, pp. 367-422, 1982.

59. Blackmore, D.R., Herman, M.N., and Woodward, J.L. "Heavy Gas Dispersion Models." J. Haz. Mat., 6, pp. 107-128, 1982.

60. Barad, M.L. (Ed.) "Project Prairie Grass, A Field Program in Diffusion." Geophysical Research Papers, No. 59, Vols. I and II. Air Force Cambridge Research Center, 1958.

61. Barad, M.L., and Fuquay, J.J., (Eds.) "The Green Glow Diffusion Program." Geophysical Research Papers, No. 73, Vols. I and II. USAEC Report HW-71400, Air Force Cambridge Research Laboratories, 1962.

62. Taylor, J.H. "Project Sand Storm, An Experimental Program in Atmospheric Diffusion." Environmental Research Papers, No. 134, Report AFCRL-65-649, Air Force Cambridge Research Laboratories, Cambridge, MA, 1965.

63. Agterberg, R., Nieuwstadt, F.T.M., Van Duuren, H., Hasselton, A.J. and Krijt, G.D. Dispersion Experiments with SF_6 from the 213m High Meteorological Mast at Cabauw in the Netherlands, KNMI, 1983.

64. Driedonks, A.G.M., van Dop, H., and Kohsiek, W.H. "Meteorological Observations on the 213 m Mast at Cabauw, in the Netherlands. Fourth Symposium on Meteorological Observations and Instrumentation, Denver, April 10-14, American Meteorological Society, 1978.

65. Nieuwstadt, F.T.M., and van Duuren, H. "Dispersion Experiments with SF_6 from the 213 m High Meteorological Mast at Cabauw in The Netherlands." Fourth Symposium on Turbulence, Diffusion, and Air Pollution, Reno, Nevada, Jan. 15-18, American Meteorological Society, 1979.

66. van Duuren, H., and Nieuwstadt, F.T.M. "Dispersion Experiments from the 213 m Meteorological Mast at Cabauw in The Netherlands." Proceedings of the 14th International Colloqium, Paris, France, May 5-8. M.M. Benarie (Ed.), Studies in Environmental Science, Volume 8, Elseviers Scientific Publishing Company, Amsterdam, 1980.

67. Hanna, S.R. "The exponential probability density function and concentration fluctuations in smoke plumes." Bound. Lay. Meteorol., 29, pp 361-375, 1984.

68. Hanna, S.R., Weil, J.C. and Paine, R.J. Plume Model Development and Evaluation: Hybrid Approach. Final Report to EPRI, 3412 Hillview Ave., Palo Alto, CA 94303, October 1986.

69. Strimaitis, D.G., G.F. Hoffnagle, and A. Bass, On-Site Meteorological Instrumentation Requirements to Characterize Diffusion from Point Sources - A Workshop, U.S. EPA, Research Triangle Park, NC 27711, 1980.

70. Kaimal, J.C., J.E. Gaynor, P.L. Finkelstein, M.E. Groves, and T.J. Lockhart, A Field Comparison of In Site Meteorological Sensors, Boulder Atmospheric Observatory Report No. 6, NOAA/ERL, Boulder, CO, 80303, 1984.

71. White, J., R. Ewald, and C. Biltoft, Site Influences on the XM76 Grid. DPG Doc. No. DPG-FR-86-701, U.S. Army Dugway Proving Ground, Dugway, UT 84022., 1986.

72. Briggs, G.A. Diffusion Estimation for Small Emissions. ATDL Contribution File No. 79, Atmospheric Turbulence and Diffusion Laboratory, 1973.

73. Luna, R.E., an H.W. Church, "A Comparison of Turbulence Intensity and Stability Ratio Measurements to Pasquill Stability Classes." Journal of Applied Meteorology, 11, pp 663-669, 1972.

74. Chatwin, P.C. "The Use of Statistics in Describing and Predicting the Effects of Dispersing Gas Clouds." Dense Gas Dispersion, (Ed. by Britter and Griffiths), Elsevier, New York, NY, pp. 213-230, 1982.

75. Meroney, R.N., and Lohmeyer, A. "Prediction of Propane Cloud Dispersion by a Wind-Tunnel Data Calibrated Box Model." J. Haz. Mat., 8, pp. 205-221, 1984.

76. Wilson, D.J. Predicting Risk of Exposure to Peak Concentrations in Fluctuating Plumes. University of Alberta, 1982.

77. Lewellen, W.S., and Sykes, R.I. "Analysis of Concentration Fluctuations from Lidar Observations of Atmospheric Plumes." J. Clim. and Applied Meteorol., 25, pp. 1145-1154, 1986.

78. Rafferty, J.E., and Dumbauld, R.K. Analysis of Transmissometer Data for Smoke Grenades Tested During the Comparison Test of Grenade, Smoke: Screening, RP, LA81. Dugway Proving Ground Technical Report TR-83-342-01, 1983.

79. Golder, D. "Relations Among Stability Parameters in the Surface Layer." Boundary Layer Meteorol., 3, pp 47-58, 1972.

80. Fox, D.G. "Judging air quality model performance - review of the Woods Hole Workshop." Bull. Am. Meteorol. Soc., 62, pp 599-602, 1981.

81. Environmental Protection Agency, Interim Procedures for Evaluating Air Quality Models (Revised), EPA-450/4-84-023. OAQPS, EPA, Research Triangle Park, NC., 1984.

82. Cox, W.M. and Tikvart, J.A. "Assessing the performance level of air quality models." Paper V. 3 at 15th Int. Tech. Meeting on Air Pollution Modeling and its Applications. NATO/CCMS, 1985.

83. Hanna, S.R., and Heinold, D.W. Development and Application of a Simple Method for Evaluating Air Quality Models. API Publ. NO. 4409, Amer. Petroleum Inst., 1220 L St., N.W., Washington, DC, 20005, 38 pp., 1985.

84. Weil, J.C., and Brower, R.P. "An Updated Gaussian Plume Model for Tall Stacks." J. Air Poll. Control Assoc., 34, pp. 818-827, 1984.

85. Efron, B. The Jacknife, The Bootstrap, and other Resampling Plans, CBMMS-NSF-38. Soc. Ind. and Appl. Math., Philadelphia, 92 pp, 1982.

86. Panofsky, H.A., and Brier, G.W. Some Applications of Statistics to Meteorology. The Pennsylvania State University, 224 pp., 1968.

87. Tukey, J.W. Kinds of Bootstraps and Kinds of Jackknives, Discussed in Terms of a Year of Weather Related Data. Tech. Report No. 292. Princeton, Univer., 1987.

88. Alp, E., Oliverio, M., Lam, L.H., Matthias, C.S., and Bourque, C.P., COBRA: A Heavy Gas/Liquid Spill and Dispersion Modeling System, 1987.

(The reverse of this page is blank)

89. Hanna, S.R., "The Effect of Line Averaging on Concentration Fluctuations." Bound. Layer Meteorol., 40, pp 329-338, 1987.

(The reverse of this page is blank.)

APPENDIX A

DETAILED DISCUSSION OF FIELD EXPERIMENTS

APPENDIX A - DETAILED DISCUSSION OF FIELD EXPERIMENTS

A. PROJECT PRAIRIE GRASS

1. Overview

Project Prairie Grass was held in north-central Nebraska near O'Neil in the summer of 1956 (Reference 60). Personnel from the Massachusetts Institute of Technology, the Texas A. and M. Research Foundation, the University of Washington, the University of Wisconsin, the Air Weather Services, and the Air Force Cambridge Research Center participated in this series of 70 trials. The project was designed by Air Force personnel at the Air Force Cambridge Research Center. The primary objective was to determine the rate of diffusion of the continuously emitted SO_2 tracer gas as a function of meteorological conditions. Secondarily, it was hoped that insight would be gained into turbulence phenomena. Releases were of 10 minutes duration and from ground level. Measurements were made at 50, 100, 200, 400, and 800 meters downwind. The trials were made over flat prairie terrain under a variety of meteorological conditions. Approximately half of the trials were conducted during unstable daytime periods and the rest were held at night with temperature inversions present.

2. Site Description

The experiment site was located about 5 miles northwest of the town of O'Neil in Holt County, Nebraska (latitude 42 degrees, 29.6 minutes north; longitude 98 degrees, 34.3 minutes west). Approximately 1 square mile, designated Section 14, Township 29 North, Range 11 West, was leased for the duration of the experiment. The site was virtually level and covered with natural prairie grasses, which were mowed prior to the start of the tests and grew very little over the period. There was an unobstructed view for miles with no distinct horizon visible except to the southeast where a small hill was located. A-C power was available from lines running approximately perpendicular to the array centerline at a distance from the release point of about 1 kilometer. The site was nearly completely obstruction free. The nearest farmhouse was well over 1300 meters northwest fo the release point; no complaints were made about the gas by any nonparticipants. The sampling grid was located on semicircular arcs at the aforementioned downwind distances at bearings between 180 and 270 degrees. This array setup was chosen primarily because the wind climatology indicated that the wind was from between 120 and

240 degrees more than 50 percent of the time in July and August. The soil was composed of a black top soil about 25 centimeters thick, a brown subsoil about 20 centimeters thick, and, beneath this brown subsoil, a light brown layer of compacted soil about 15 centimeters in depth. The top soil contained about 4 percent organic material. Both the top soil and the subsoil had good water-holding capacity. Beneath the compacted layer was a loose, coarse sand of about 60 centimeters depth. Water held in the sandy layer affected the surface vegetation only very slowly since very few roots penetrated so far and upward water movement was quite slow through the sand and compacted soil.

3. Experimental Design

The project was designed with several points in mind, notably the improved understanding of general turbulence theories, the testing of specific diffusion theories, and an attempt to experimentally verify both past and (then) present theories. These points helped determine what types of meteorological measurements were necessary. It was hoped that related problems ranging from crop dusting to the forecasting of low level wind shear would be brought nearer to solution by the results of the project. The tracer technique used was designed by M.I.T. at its Round Hill Field Station. The technique involved the continuous emission of sulfur dioxide. A continuous source was chosen for the following reasons: first, continuous sources are more easily reproduced for projects comprised of many trials, second, interpretation of the concentration data is somewhat easier, and third, the question of what meteorological data is pertinent is usually simpler. The duration of the releases was 10 minutes. This was chosen after consideration of factors such as the cost of the tracer gas, distance between samplers, and practical rates of emission. It was highly desirable that tracer losses on vegetation and the ground be negligible, at least in the sampling area. Another necessity was that the sampler analysis be accurate, cover a wide range of concentrations, and be done in a timely fashion.

4. Release Technique

A continuous point source of sulfur dioxide located near ground level was used in Project Prairie Grass. SO_2 was chosen as the tracer in part because it was relatively inexpensive and easily available. Liquid SO_2 was vaporized in a specially constructed chamber immersed in hot water in a large circular tank. The water was maintained at a temperature of 50 degrees

Celsius; it provided the needed thermal energy to insure that the emission rate remained constant (the rate would otherwise drop off as the pressure fell due to the rapid cooling that occurred during the phase change). The gas then flowed through a pressure regulator and an adjustable flow-controller valve. The pressure and temperature of the gas were measured at the inlet and the outlet to aid in an accurate determination of the source strength. This apparatus was partly buried in a trench to reduce its effects on the natural air flow. The tracer was emitted horizontally from a 2 inch plastic pipe at a height of 46 centimeters. During the daytime releases, the maximum source strength of 100 grams per second was used. The emission rate varied less than 5 percent in almost all of the releases.

5. Sampling

Midget impingers from the Mine Safety Appliance Company were used to make the measurements. Each impinger contained 10 milliliters of dilute hydrogen peroxide solution. Air drawn into the impingers via aspiration by vacuum units was broken into very small bubbles as it impacted upon the bottom of the glass flasks. SO_2 in the air would react with the hydrogen peroxide to form sulfuric acid. The impingers were mounted on steel fence posts at a height of 1.5 meters along 5 semicircular arcs. The steel posts were at 2 degree intervals along the 50, 100, 200, and 400 meter downwind arcs and at 1 degree intervals along the 800 meter arc. Average SO_2 concentrations were also determined in the vertical at the 100 meter arc. Lightweight television-type towers were spaced at 14 degree intervals and instrumented at 9 levels: 0.5, 1, 1.5, 2.5, 4.5, 7.5, 10.5, 13.5, and 17.5 meters. The analysis of the collected impingers was accomplished by measuring the electrical conductance of the aspirated solutions using Wheatstone bridges and dipping conductivity cells. Each impinger was placed in a constant temperature water bath. When the temperature stabilized at the appropriate value, the conductivity cell was placed in the impinger and the resistance indicated on the Wheatstone bridge was recorded. The conductance values were converted to gas concentrations using well known laboratory techniques. The uncertainty involved in determining the conductance was less than 2 percent in the normal range of concentrations.

6. Meteorological Data

Cup anemometers and airfoil type vanes were used to measure wind speed, direction, and fluctuations in wind direction at a height of 2 meters

at 2 locations: one pair of instruments at the release point (one to each side of the source and about 25 meters away from it) and another pair 450 meters downwind about 30 meters west of the centerline. The two cup anemometers used were chosen from a group of 11 similar devices on the basis of field-matching tests - an average difference of only about 0.25 percent in calibration was found. Wind speed data was recorded on Esterline-Angus chart recorders. The balsa wood airfoil vanes used to measure wind direction and direction fluctuations became deformed due to exposure to the wind and the rain. These devices were replaced after trial 34 with vanes made with flat metal plates. The instrumentation was operated for 20 minute sampling periods that were centered on the midpoint of the 10-minute gas releases. Mean wind speed, direction, and standard deviations of direction were calculated both for the 20 minute sampling periods and the 10-minute release periods. The wind speed data was thought to be accurate to within 2 to 5 percent for mean wind speeds greater than 2 meters per second. The uncertainty was significantly larger for lower mean wind speeds, such as would occur at night under stable atmospheric conditions (the starting speed of the cup anemometers was 0.8 meters per second). The wind direction values may be in error by up to 10 degrees. Standard deviations of the wind direction were considered accurate to within 10 percent except when the mean wind speed was less than 2 meters per second.

The Texas A and M group had a variety of instrumentation located about 825 meters downwind of the source and 225 meters west of the array centerline. These instruments included air sampling tubes, thermocouples, soil temperature elements, cup anemometers and wind vanes, a pyrheliometer to measure incoming shortwave radiation, and a net radiometer. Wind vanes were located at heights of 1 and 16 meters, anemometers were sited at heights of 0.25, 0.5, 1, 2, 4, 8, and 16 meters, thermocouples and air samplers were placed at heights of 0.125, 0.25, 0.5, 1, 2, 4, 8, and 16 meters, and the soil temperature elements were placed at 0.031225, 0.0625, 0.125, 0.25, 0.5, and 1 meter below the surface. The measurements made by these devices were used to evaluate the latent and sensible heat fluxes. Rawinsonde flights were made by the 6th Weather Squadron (Mobile), Tinker Air Force Base, Oklahoma. Flights were made for all trials except 35 S and 48 S. Computations were made according to standard Air Weather Service procedures. Pressure, height, temperature, and relative humidity values were tabulated for the significant and mandatory levels. Wind values were included for the standard flights. Aircraft soundings were also taken at the times of the diffusion trials. The

data consisted of height, pressure, temperature, relative humidity, vapor pressure, and dew point temperature. A standard U.S. Air Force L-20 that had been instrumented at Hanscom Air Force Base, Bedford, Massachusetts, was used. Flights were made for all trials except 23, 24, 31, 32, 33, and 34.

B. PROJECT GREEN GLOW

1. Overview

The Green Glow series of tests were so named because the zinc sulfide particles used as a tracer give off a green fluorescence under ultraviolet light sources (Reference 61). The primary objective of the field study was to calculate, as functions of the meteorological conditions, the horizontal and vertical diffusion rates of the particulate tracer. Measurements were to be made by a dense sampling grid over as great a distance as possible (it was hoped out to 16 miles). Green Glow was, in some ways, an extension of Project Prairie Grass. Measurements were made at the same heights above ground level, 1.5 meters, and at 2 of the downwind distances used in Prairie Grass, 200 and 800 meters. Other measurement distances were 1600, 3200, 12800, and 25600 meters. 26 trials were done on the Hanford reservation of the U.S. Atomic Energy Commission near Richland, Washington in 1959. All trials were conducted at night over slightly rolling terrain. In addition to horizontal measurements of the tracer, vertical distributions were also measured at the first 4 arcs. Green Glow was a joint program designed by personnel at the Hanford Laboratories of General Electric and the Geophysics Research Directorate of the Air Force Cambridge Research Laboratories. Site preparation, equipment procurement and installation, making the measurements, and reducing the data were the responsibility of the General Electric personnel. Air Force personnel helped coordinate the efforts of the various participants and aided in the measurement and data reduction phases.

2. Site Description

The Hanford reservation is located in south-central Washington, approximately 30 miles east of Yakima and 125 miles southwest of Spokane. The reservation is surrounded on all sides by elevated terrain, varying from 3500 feet on its southern border with the Rattlesnake Hills to about 1100 feet on the eastern rim of the basin. There are several major breeches in the basin sides - the Beverly Gap on the northwest side, the Ringold Coulee on the east

side, the broad valley to the southeast, and to the south, the Benton City Gap. These openings channel the air flow into and out of the basin and also lead to mountain-valley circulations. The topography usually causes a drainage flow over the central part of the Hanford reservatiion from the northwest to the southeast. The sampling grid was located on the valley floor with the baseline approximately parallel to the major ridges. Vegetation consisted of desert grasses interspersed with sagebrush 1 to 2 meters in height. The locale was quite flat, dropping only 300 feet over 16 miles.

3. Experimental Design

Planning was guided by the decision to tie in Project Green Glow with Project Prairie Grass, by the peculiarities of the Hanford Fluorescent Tracer System, by the topography of the site, and by the economics and logistics of the situation. These factors gave rise to the following guidelines:

a. Sampling arcs would be concentric about the source.

b. Two arcs would be 200 and 800 meters from the source.

c. The release would be from ground level.

d. The maximum release rate would be 8 kilograms per hour.

e. The counting-statistics-based assaying system used would require a minimum of about 100 particles per sample.

f. The range of the sample-assaying system was only 5 orders of magnitude.

g. To provide the necessary accuracy in arcwise dispersion estimates, the centerline dosages would have to be at least 100 times the minimum significant count.

h. The tracer would deposit on the surface and or vegetation at an unknown rate.

i. Releases would be of 30 minutes duration.

j. Dosage rather than average concentration would be measured.

k. All trials would be at night under stable atmospheric conditions.

l. Samples would be collected at a height of 1.5 meters.

The major design tasks were to specify the geometry of the sampling network, the sampler spacing, and the rates of sampling. These parameters had to acknowledge the constraints imposed by the aforementioned guidelines. It must be noted that the design remained flexible throughout the project so that encountered problems could be dealt with quickly; errors or oversights were caught by performing a continual qualitative check of the data.

4. Release Technique

Two standard Todd Insecticidal Fog Applicators dispensed the tracer. These devices are aerosol fog generators that consist of an air blower, a combustion chamber, a formulation pump, and a gasoline engine. The formulation used consisted of zinc sulfide pigment mixed with sodium lauryl sulfate (a surface active agent) and a small amount of water. Special care was taken to insure that the pigment was of a uniform size distribution. Each fog generator output about 20 grams per hour and this rate was more or less constant.

5. Sampling

The samplers used in Green Glow consisted of a membrane filter contained by a disposable polyethylene filter holder. They were bulk samplers, intended to collect all particles in the intake zone. A new set of sampling units was used for each trial. The bulk samplers were assayed by different means, depending on whether or not there was a significant amount of dust on the filter. If the filters appeared relatively free from dust, assaying was accomplished by use of a Rankin counter. Rankin counters use a radioactive isotope (plutonium) to activate the fluorescent pigments on the filters via alpha bombardment. The scintillations were viewed and counted by a multiplier phototube. Background counts on the Rankin devices were fairly low - between 2 and 8 counts per minute. In cases where dust contamination was present, a Tri-Carb liquid scintillation spectrometer was used. The samples were treated with a solvent composed of 3 parts ethyl acetate and 1 part ethyl alcohol. As nearly all the airborne particles were insoluble,

including the tracer, a suspension of tracer and contaminants was produced after agitation. The sample was irradiated by the spectrometer and then the phosphorescence was counted with a multiplier phototube. To reduce the background counting level, the phototube system was placed in a deep freeze.

6. Meteorological Data

Meteorological data was collected on the instrumented 410 foot tower, on a portable 78 foot mast, from the Hanford radio-telemetering network, and from rawinsonde launches made by personnel from the 6th Weather Squadron (Mobile), Tinker Air Force Base, Oklahoma. Measurements on the 410 foot tower were made at heights of 3, 7, 50, 100, 150, 200, 250, 300, and 400 feet. Wind speed and direction were measured at all levels except at 3 feet, temperature at all levels except 7 feet, and dew point temperature at all levels except 7, 150, and 250 feet. Aspirated copper thermo hms were used to measure temperature, Foxboro Dew Cells were used to measure dew point temperature, and Friez Aerovanes were used to measure the wind velocity. All information was recorded on strip charts. Wind speed and temperature were measured at 2.5, 5, 10, 20, 40, and approximately 80 feet on the portable mast provided by General Electric. Wind direction was measured at 2.5, 10, 40, and 80 feet using Beckman and Whitley indicators. Wind speed was measured by 3 cup anemometers from C.F. Casella and Company, Limited, that had been modified into photo-chopping devices. Thermocouples were used to measure temperature. All signals were recorded on a single strip chart recorder using a scanning unit from Pannelit, Inc. Wind speed and direction were measured at the 18 remote stations comprising the Hanford radio-telemetering network. The central receiving station was located at the 410 foot tower, near the release point. Modified Friez Aerovanes placed at at height of 23 feet were used for the measurements. Standard Air Weather Service equipment was used by the team from the 6th Weather Squadron to make the rawinsonde observations. Ascents were made 1 hour in advance of each release time, at the start of each release, and 1 hour after each release time. All computations were done by the Weather Squadron personnel. The 1 minute angle values from the wind recorder were used in preparing the data contained in the project report.

C. PROJECT OCEAN BREEZE

1. Overview

Project Ocean Breeze was conducted at Cape Canaveral, Florida, during 1961 and 1962 (Reference 27). Air Force personnel, General Electric technical staff, and various Air Force contractors participated in the setup and or operation of the 76 trials that comprised the project. The trials were conducted both in the summer and winter seasons; sea breezes occurred frequently in the summer, and numerous passages of cold fronts in the winter brought about unstable conditions. The primary objective was to provide the data needed to develop and test a set of diffusion prediction equations to be used operationally at the Cape Canaveral missile test range. Air Force Cambridge Research Laboratory was also determined to develop and install an automated meteorological data acquisition and processing system on base. This system would continually output solutions to the developed diffusion prediction equations. The Hanford Tracer System was used in the project; its features were factored into the overall project design. Zinc sulfide particles were used as the tracer. The release point was placed between launch pads 15 and 16, approximately 1000 feet from the ocean. Measurements were made at a height of 15 feet at concentric arcs located 0.75, 1.5, and 3.0 miles downwind of the source. The bulk-collecting membrane filters were placed at 15 feet above ground level because of the vegetation, which was composed of palmetto 2-5 feet tall and brushwood 7-14 feet tall, situated on rolling sand dunes 10-20 feet in height.

2. Site Description

The Cape Canaveral missile range is located on the east-central Florida coast. Its eastern side is bordered by the Atlantic Ocean, and its western border is the Banana River. The experiment site was characterized by 10-20 feet tall rolling sand dunes. In addition, much of the diffusion course was covered with brushwood and palmetto growth. The sampling grid was located on the aforementioned arcs at 2 degree intervals on arcs 1 and 2 between 152 and 340 degrees with respect to the source, and at 1.5 degree intervals on arc 3 between 152 and 236.5 degrees. The orientation of arc 3 limited its use to occurrences of northerly winds, which occurred fairly often during the winter months.

3. Experimental Design

The experiment was designed to be of use in determining the potential effects of accidental toxic releases on the missile range. As such, the most probable accident scenarios were considered and found to be able to be characterized by a continuous point source located at ground level. Due to the large variability in possible emission periods and to the cost factor, releases were designed to be of 30 minutes duration. The parameters used to decide upon the geometry of the sampling grid were climatological data for the site, terrain features, and the experience of Project Green Glow. Some of the initial criteria used in the design process were:

a. Sampling would be done along logarithmically spaced arcs located at distances determined by the scale of the problem.

b. The release would be from ground level and of 30 minutes duration.

c. The maximum release rate would be 8 kilograms per hour.

d. The sample assaying would be done at the Hanford Plant.

e. Dosages rather than concentrations of the tracer would be measured.

4. Release Technique

As in Project Green Glow, the zinc sulfide tracer was released using two standard Todd Insecticidal Fog Applicators. These aerosol fog generators consist of an air blower, a combustion chamber, a formulation pump, and a gasoline engine. The formulation used was composed of zinc sulfide pigment, the surface active agent sodium lauryl sulfate, and water. Tracer emission rates were calculated by subtracting formulation amounts in the holding tank at the end of each run from the amount recorded prior to each run. The effective source height was 2-3 meters above ground level.

5. Sampling

The samplers used were the same as those used in Project Green Glow - membrane filters in disposable polyethylene holders. These bulk samplers were

used only once; there was a new set of samplers for every trial. The samplers were assayed using a Rankin counter. This device activates the fluorescent particles embedded in the filter using alpha bombardment brought about by a 200 microcurie plutonium source. The scintillations that occur are viewed by a multiplier phototube, amplified, and recorded using a scaler. The background counting rate was about 5 counts per minute, which was equivalent to approximately 4 E-9 grams of tracer. Unlike Project Green Glow, dust contamination was not a problem.

6. Meteorological Data

Meteorological measurements were made by personnel at Pan-American's Cape Weather Station, located a few hundred feet downwind from arc 2 at a bearing of 196 degrees from the source. These measurements consisted of wind speed and direction data using a Belfort Type M device sited at a height of 12 feet above ground level and temperature profiles from wiresonde captive instrumented balloons. Standard synoptic and rawinsonde data were provided by Detachment 11, 4th Weather Group, Air Weather Service, Patrick Air Force Base. The Weather Detachment also made the necessary wind direction forecasts used in the scheduling of each diffusion trial.

D. PROJECT DRY GULCH

1. Overview

Project Dry Gulch was conducted at Vandenburg Air Force Base, California, during 1961 and 1962 (Reference 27). U.S. Air Force personnel, General Electric personnel from the Hanford complex in Washington, and various contractors aided in the setup and or operation of the 109 trials that comprised the project. Preparation of the diffusion course was done by the Martin Company, training of the field crews, scheduling of the trials, furnishment of the tracer and sampling filters, and tabulation of the data was accomplished by General Electric. Air Force personnel made meteorological measurements, and rawinsonde data was made available by the U.S. Weather Bureau station at Point Arguello. The primary objective of Dry Gulch was the same as that of Project Ocean Breeze - the development and testing of diffusion prediction equations from the experimental data. These equations were needed for use in situations of accidental releases of toxic gases at the missile range. The Hanford Tracer System was used at Vandenberg. This system used fluorescent zinc sulfide particles dispensed by aerosol fog generators as

the tracer. Measurements were made on two diffusion courses - course B was located on the mesa that was 200-300 feet in elevation with a 40-60 foot bluff at the coastline, course D was located in the Lompoc Valley which runs from west-northwest to east-southeast along the southern edge of the aforementioned mesa. All samples were placed at a height of 1.5 meters above ground level. The source point for course B was about 2600 yards inland; the source point for course D was approximately 1100 yards from the coastline. Course B had 2 concentric sampling arcs which were 1.43 (2301 meters) and 3.52 (5665 meters) miles downwind. Course D had 3 arcs located 0.53 (853 meters), 0.93 (1500 meters), and 2.93 (4715 meters) miles downwind.

2. Site Description

The terrain at the experiment site was quite complex. It consisted of a broad, 200-300 feet tall mesa which sloped westward and terminated in a 40-60 foot bluff at the coastline. Several sharp ravines cut the mesa and 2 valleys, running roughly west-northwest to east-southeast, denoted the northern and southern termini of the mesa. The southern valley, the Lompoc, is quite broad with gently sloping sides; the San Antonio Valley to the north of Burton Mesa is fairly narrow with relatively steep sides. The samplers were placed at 2 degree intervals on arc B-1 from 87 to 171 degrees with respect to the source, and at 1 degree intervals on arc B-2 from 85 to 171 degrees. Samplers were located every 2 degrees from 60 to 180 degrees on arcs D-1 and D-2, and every 1 degree from 110 to 180 degrees on arc D-3.

3. Experimental Design

As stated in the overview, Project Dry Gulch was designed to aid in the determination of the potential effects of accidental releases of toxics on the missile range. The most probable accident types were found to be capable of being modeled by a ground level, continuous point emission. Since actual emission times may vary significantly, all releases were designed to be of 30 minutes duration. There was a great deal on interest in sea breeze conditions and the resultant diffusion because of the high frequency of occurrence of such situations. The geometry of the sampling grid was formulated using climotological data for the site, terrain features, and the experience of similar tests in Project Green Glow. Some of the criteria used to design the tests were:

a. Sampling would be done along logarithmically spaced arcs located at distances determined by the scale of the problems considered.

b. The release would be from ground level.

c. The release would last for 30 minutes.

d. The maximum release rate would be 8 kilograms per hour.

e. All sample assaying would be done at the Hanford Plant.

f. Dosages rather than time averaged concentrations of the tracer would be measured.

4. Release Technique

Zinc sulfide was released using 2 Todd Insecticidal Fog Applicators. These devices are aerosol fog generators that consist of an air blower, a combustion chamber, a formulation pump, and a gasoline engine. The actual formulation used in the fog generators was composed of zinc sulfide pigment, the surface active agent sodium lauryl sulfate, and a small amount of water. The effective source height was 2 to 3 meters above ground level. Tracer emission rates were calculated from the measured amounts of the formulation recorded before and after each trial.

5. Sampling

Membrane filters fitted into disposable polyethylene holders were used as the sampling devices. The filters were of the same type as those used in Project Green Glow. As in that previous experiment, these bulk samplers were used only once; a new set of samplers was required for each trial. The samples were assayed with a Rankin counter and a multiplier phototube. The counter was used to activate the embedded fluorescent particles via alpha bombardment. The excitation caused scintillations that were dutifully viewed by the phototube then amplified and recorded by a scaler device. It was thought that the background counting rate of 5 counts per minute was low enough to be able to distinguish a goodly range of tracer concentration.

6. Meteorological Data

Members of Detachment 3, 3rd Weather Wing, Air Weather Service, provided supporting meteorological measurements. Belfort Type M devices placed 12 feet above ground level were used to measure wind speed and direction. For the first 29 trials, measurements of temperature differences (delta-T) over certain heights were made using a "gantrysonde". This not terribly trustworthy device was composed of wiresonde instruments mounted on an Atlas rocket gantry. Wiresondes replaced the "gantrysonde" after trial number 29. The U.S. Weather Bureau rawinsonde station at Point Arguello made many special launches and detailed calculations of the temperature and wind profiles up the 700 millibar pressure level. While this data was certainly of use, it should be noted that the rawinsonde launch site was 8 miles south of and 200 feet higher than the wiresonde site.

E. THORNEY ISLAND

1. Overview

Thorney Island, England, was the site of 16 unobstructed, large-scale releases of a heavy gas tracer during the summers of 1982 and 1983 (Reference 47). The experiment was conducted by the National Maritime Institute under contract to the United Kingdom Health and Safety Executive with the sponsorship of numerous international organizations. The main objective was to measure and archive data for the sponsoring organizations to use in the analysis and testing of the capabilities of various computer models. To this end, much effort went into insuring that there would be as few ambiguities as possible associated with the data. For example, the mechanism for releasing the gas tracer provided well defined initial conditions for model calculations. Instantaneous releases of 2000 cubic meters of a gas mixture of Freon-12 and nitrogen were accomplished using an accordion-type container. Relative densities with respect to air were between 1.5 and 4. Many meteorological parameters were measured, including wind speed and direction, turbulence, temperature, humidity, and solar radiation. Concentrations of the dispersed tracer were measured using gas sensors, most with a frequency response of 1 Hertz, that actually recorded oxygen deficiency. Thirty-eight fixed masts, instrumented at 0.4, 2.4, 6, and 10 meters, were placed in a 100 by 100 meter grid. The gas sensors were actually placed at heights of 0.4, 2.4, 4.4, and 6 meters above ground level. The wind data was measured at 10 meters. The gas measurements were averaged over 0.6 second intervals in an

attempt to filter out sensor response noise and to retain the shape of the sensor responses.

2. Site Description

Thorney Island is located near latitude 51 degrees, longitude 1 degree. It is the site of an abandoned airfield; the runways are still extant. The experiment site was flat and uniform over an area of 1 by 0.5 kilometers. Upwind of the sampling grid, the fetch was clear for 1 kilometer. The surface was grass interspersed with tarmac runways. The grass was approximately 15 centimeters in height and the roughness length determined for the locale was 1 centimeter. In an attempt to minimize temperature differences between the grass and the runways, all portions of the runway within the sampling grid were painted white.

3. Experimental Design

The experiment was designed to provide better understanding of the hazards involved in the storage and transportation of heavier-than-air gases. Knowledgeable judgement of the risk involved in the use of materials such as ammonia and liquified natural gas requires a great deal of information. Prominent in the list of such necessary data is the diffusion of heavier-than-air gases. The Thorney Island study was specifically designed to provide data on the dispersion of large-scale instantaneous releases of negatively-buoyant gases and to use this data to quantitatively determine the capabilities of various classes of computer models. Thorney Island was chosen as the test site primarily because it met the meteorological, topographical, and logistical requirements (wind direction steady for goodly periods of time or for predictable intervals based on the synoptic and or mesoscale weather conditions, flat and unobstructed, easy to supply).

4. Release Technique

Tracer production was carried out with gas fired vaporizers. The liquid Freon-12 and nitrogen were vaporized then mixed in the gaseous phase and finally pumped to the release site. The release device consisted of an inflated, flexible fabric cylinder 14 meters in diameter and 13 meters in height. This fabric bag was attached to horizontal, metal rings, each of which was supported by cables that in turn were attached to a 22 meter tall central tower. The cables were disconnected at release times, causing the

fabric bag to fall concertina-fashion. The collapse of the bag was complete in 1 to 2 seconds, leaving a free-standing cylinder of gas. Orange smoke mixed with the gas acted as a marker for the photographic records of the trials. The release volume was determined from the size of the fabric bag and the meter readings of the individual gas storage tanks.

5. Sampling

The gas monitoring devices sited at 0.4, 2.4, 4.4, and 6 meters on each of the 38 fixed and 4 mobile masts measured oxygen deficiency. The level of oxygen deficiency was related to the amount of tracer present in a known manner. One hundred and eighty sensors had a 1 Hertz frequency response, 5 sensors were faster in response (10 Hertz). Data was acquired using 35 conversion units, each of which provided analog to digital conversion for 8 data channels. Each of the conversion units was connected to a central computer which collected and archived the digitized data. After the archiving of the data, the voltage measurements were converted to engineering units using the appropriate scales. Once this was accomplished, the gas sensor data validated to check the gain and zero shift of the instruments. The data was then made available to the sponsors.

6. Meteorological Data

There were a total of 32 meteorological instruments in use during the Thorney Island trials. Ten of these devices were three-dimensional anemometers. Quantities measured included wind speed and direction, temperature, pressure, relative humidity, and solar insolation. Several methods were used to determine the Pasquill-Gifford stability class and they produced a variety of values. The methods used were delta-T, the indirect method of determining the sensible heat flux proposed by Smith in 1979 using solarimeter data, direct measurement of the sensible heat flux, Richardson number, bulk Richardson number, and sigma theta. Of this plethora of methods, the sigma theta scheme seemed to work best. The delta-T method gave a result of class A for each of the unobstructed trials (numbers 6-19). The direct and indirect heat flux methods always resulted in class F. With such variation between methods, it required careful consideration of the meteorological conditions to ascertain the atmospheric stability.

(The reverse of this page is blank.)